THREE ESSAYS ON PREDICTING THE EQUITY PREMIUM AND ASSET RETURNS

THREE ESSAYS ON PREDICTING THE EQUITY PREMIUM AND ASSET RETURNS

By ADAM STIVERS

B.S., M.A.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the

Requirements for the Degree Doctor of Philosophy

McMaster University

Hamilton, Ontario

DOCTOR OF PHILOSOPHY (Business Administration – Finance), 2016

TITLE: Three Essays on Predicting the Equity Premium and Asset Returns

AUTHOR: Adam Stivers,

B.S., M.A. (West Virginia University)

SUPERVISOR: Professor Ron Balvers

NUMBER OF PAGES: x, 166

Abstract

This dissertation contains three essays on the predictability of asset returns and its implications. All essays contain tests and applications that could be implemented in real time, and a summary of economic and statistical significance. The dissertation as a whole adds to the growing literature that returns can be predictable, enough so to provide economically significant differences in returns. The same applies to predicting market returns, or market timing. The first essay modifies a popular measure of investor sentiment (Baker and Wurgler, 2006) with the idea in mind that market-wide sentiment will partially reverse itself in the next period. Economic fundamentals are removed to provide a more pure measure of sentiment, which is found to predict returns better (with sentiment having a negative relationship with market returns), resulting in a better investment strategy. The second essay implements a statistical methodology which allows many predictors (portfolio returns) to be aggregated into a composite index. This index predicts returns well, and also gives insight into why two well-observed stock market anomalies, size and value premiums, may occur: they predict future market returns and are therefore an ICAPM state variable that reflects future wealth opportunities. For this reason they carry a risk premium. The third essay provides a new forecasting approach that imposes the restrictions of the popular arbitrage pricing theory (APT) on an existing statistical approach (principal components analysis). The result is that the expected returns of asset positions

hedged against systematic risk are better estimated, and average error is greatly reduced out-of-sample.

Acknowledgements

I would first and foremost like to thank my supervisor, Professor Ron Balvers, for his immense help with this dissertation and developing my research and writing skills. I have learned a great deal from him about finance and academic research in general. I thank my other dissertation committee members, Professors Richard Deaves and Rosemary Luo. They have provided valuable comments and suggestions over the years, as well as being a valuable resource for me approach with questions. I would also like to thank my external examiner, Professor Ilias Tsiakas, for his useful comments and suggestions as well as his time. I would also like to thank my fellow graduate students and all those who attended the in-house seminars at which I presented my research at various stages. This includes Professors Clarence Kwan, John Maheu, Peter Miu, Jiaping Qiu, Narat Charupat, Trevor Chamberlain, and Anna Danielova. I also appreciate the support of other staff members here at McMaster, such as our program and department administrators, as well as those at the Graduate Studies office.

I would like to thank the anonymous referees and editors at the *International Journal of Financial Studies*, participants at the European Financial Management Association 2015 Annual Conference, and anonymous referees from other journals and conferences.

Contents

Chapter One: Introduction 1

References 11

Chapter Two: *Forecasting Returns with Fundamentals-Removed Investor Sentiment* 13

2.1 Introduction 13

2.2 Literature Review 19

2.3 Data and Methodology 22

2.4 Results 29

2.4.1 Forecast Results 29

2.4.2 Market Timing Tests 34

2.4.3 Statistical Significance 37

2.4.4 Robustness 38

2.5 Conclusions 43

References 45

Chapter Three: *Equity Premium Predictions and Optimal Hedging with Many Predictors Using Partial Least Squares* 59

3.1 Introduction 59

3.2 Literature Review 65

3.3 Partial Least Squares Methodology 68

3.4 Data and Methodology 71

3.5 Forecast Results and Portfolio Weights Analysis 73

3.5.1 Forecast Results 73

3.5.2 Raw Portfolio Weight Analysis 78

3.5.3 Principal component weight analysis...................................82

3.5.4 PLS Factor Out-of-Sample Performance 83
3.5.5 Hedging with PLS weights 86
3.5.6 Robustness Tests 90
3.6 Conclusion 93
References 97
Chapter Four: *Arbitrage Pricing Restrictions and the Predictability of Stock Returns by Statistical Factor Analysis* 110
4.1 Introduction 110
4.2 Basic Properties of Statistical Factor Analysis 121
4.2.1 Returns Specification with Factor Structure 121
4.2.2 Factors Chosen to Minimize Errors and PCA.....................122
4.3 Incorporating Arbitrage Pricing Restriction 125
4.3.1 Adding APT Constraint 125
4.3.2 Imposing OLS Betas 128
4.3.3 Sub-Optimal Direct Restriction on the Alphas 129
4.3.4 Correlation Matrix Instead of Covariance Matrix 130
4.4 Application to Industry Portfolios 131
4.4.1 Number of Factors and MSE Breakdown 131
4.4.2 Expanding Window 133
4.4.3 Bias and Random Error..136
4.4.4 Varying time horizon 137
4.4.5 Rolling Window...141
4.4.6 Allowing for In-Sample Betas 142
4.5 Conclusion 143
References 146
Chapter Five: Conclusion 164

List of Tables

Table 2.1 48
Table 2.2 49
Table 2.3 50
Table 2.4 51
Table 2.5 53
Table 2.6 54
Table 2.7 55
Table 2.A.1 58
Table 3.1 100
Table 3.2 101
Table 3.3 102
Table 3.4 103
Table 3.5 104
Table 3.6 105
Table 3.7 106
Table 3.8 107
Table 3.9 108
Table 4.1 152
Table 4.2 153
Table 4.3 155
Table 4.4 157
Table 4.5 158
Table 4.6 159

List of Figures

Figure 2.1 56
Figure 2.2 57
Figure 3.1 109

List of Appendices

Appendix 2.A 58

Appendix 4.A 160

Appendix 4.B 162

Chapter One: Introduction

The focus of this dissertation is to examine the potential predictability of returns; both the market returns and asset/portfolio returns. The findings add to the growing literature (Kelly and Pruitt, 2013; Li, Ng, and Swaminathan, 2013; Ferreira and Santa-Clara, 2011; Rapach, Strauss, and Zhou, 2010 as examples) that returns can be predictable up to a point where economically significant return differences can be obtained. This applies to market timing as well. In what follows, I examine the statistical and economic significance of the predictions, as well as various applications of the forecasts. Various predictors and methodologies are used in each essay. The first uses a modified measure of investor sentiment; the second, a statistical approach which creates a composite index of predictors, allowing a large-cross section of data with minimal measurement error; the third, imposing the arbitrage pricing theory (APT) restrictions while simultaneously maximizing covariance among test assets.

The purpose of the first essay is to create a better measure of investor sentiment, and then to investigate how well the new index compares with the existing measures when forecasting aggregate market returns (equity premium). The new index is formed by altering the Baker and Wurgler (2006) investor sentiment index in a way that better captures investor irrationality. Using sentiment to forecast returns follows the theory of DeLong et al. (1990). They show that noise traders can create mispricing which is then corrected in the next period. Thus, some assets may be predictable due to noise traders. The essay tests whether or not this will hold in the aggregate. If it does, then a sentiment measure that is pure sentiment (capturing only irrational behavior) should work better than other measures which may contain a rational element. The essay argues that existing sentiment measures fall short in this regard.

In an attempt to properly remove fundamentals from the Baker and Wurgler (BW hereafter) investor sentiment index, the first essay follows the methodology of Lemmon and Portniaguina (2006). Lemmon and Portniaguina separate the fundamental and non-fundamental components of consumer confidence (CC hereafter). The essay follows their methodology, but applies it to the Baker and Wurgler index by first regressing CC on a set of fundamentals, then regressing BW on the fitted (fundamental/rational) value from this initial regression, and finally using the residuals from this regression as a proxy for pure sentiment. Basically, the Lemmon and Portniaguina measure of fundamental consumer confidence is removed from the Baker and Wurgler index. Thus, a composite index of fundamentals based on consumer confidence is removed from Baker and Wurgler's composite index of sentiment. This composite index of fundamentals would limit any measurement errors or over-fitting of fundamentals when removing them from BW.

Instead of directly following Lemmon and Portniaguina, the essay uses BW for forecasting instead of CC, as it is a market measure of investor sentiment. Market measures of sentiment are composed of real actions by investors directly involved in the market. For surveys (especially consumer confidence surveys), a participant may feel overly optimistic but may not actually act on this optimism. Surveys are also open to problems such as dishonesty, bias, and incomplete responses. A market measure of sentiment (such as BW) will match investors' actions as it relates to their overall feeling regarding the stock market. Thus, the essay proposes that a market measure of investor sentiment will forecast market returns better than a survey-based measure.

To measure the predictive power of the new index, a linear forecasting model is used to forecast one-month-ahead market returns. A simple investment strategy is used: if the one-month-ahead predicted excess market return is higher than the historical average of all previous months in the sample at the time of forecasting, then the investor will invest in the market. Otherwise, she invests in a risk-free asset (T-bills). Essentially, the goal is to time the market. This differs from both Baker and Wurgler (2006) and Lemmon and Portniaguina (2006), where no market timing is involved. Market timing is considered here for two reasons. First, it is a practical application of investor sentiment, in that it is usable by practitioners. Since a common goal in the sentiment literature is predicting returns, market timing is a natural extension. The second reason for using a market timing approach is that it provides economic significance (via realized returns) in addition to statistical significance.

Following the investment strategy (essentially switching between stocks and bonds), the forecast model that removes the fundamental component of CC from BW provides average realized excess returns of around 10.6% annually compared to around 5.3% annually for BW and 8.0% annually for BWA. This realized return is found to be both statistically and economically significant. The new sentiment index also has significant market timing ability. The excess market return average over the forecasting sample is around 6.7% annually (the return from a buy-and-hold strategy), so a forecast model utilizing only BW would not beat the market on average. This evidence supports the hypothesis that investor sentiment should have no fundamental (rational) component, and that the existing measures of sentiment do not properly remove fundamentals.

The purpose of the second essay is to examine which portfolios matter more when predicting future market returns/equity premia. Specifically, portfolios sorted on size and book-to-market are used as predictors, and the methodology allows for a single composite index to be formed from a large group of predictors (portfolio returns). The results provide a possible explanation for the frequently-observed size and value premiums: these small and value stocks/portfolios carry a risk premium as they reflect changes in future wealth opportunities. Cochrane (2005) demonstrates that any variable which can explain future market returns is inherently a state variable. Hence, the composite index can be considered an ICAPM (Merton, 1973) state variable.

The essay utilizes a statistical approach, partial least squares (PLS), to allow for a large number of predictors when forecasting market returns. PLS was recently introduced to the Finance literature by Kelly and Pruitt (2013). PLS estimates jointly the factors and factor sensitivities that in-sample provide the highest covariance with the forecast variable. This also provides stable forecasts and avoids OLS issues of over-fitting, extreme weights, and collinearity. It is also shown in the essay that OLS coefficients are more volatile over time, and the out-of-sample performance of OLS is inferior to both the historical mean and PLS. PLS produces positive, diffuse, and stable weights that would be more useful in practice (in addition to superior forecasts).

To perform partial least squares, this essay uses the SIMPLS algorithm of de Jong (1993). Partial least squares can be thought of as a variant of principal-component analysis. Rather than maximizing the covariance matrix $X^T X$, PLS maximizes $Y^T XX^T Y$ or $X^T YY^T X$, which maximizes the covariance between Y and X (Y being the forecast variable(s) and X being the predictor variables in this case). Recursive forecasting is performed at a monthly frequency, with the PLS regression serving as the forecast regression. Several different Fama-French industry and size-value sorted portfolio returns are used as the predictor variables. An out-of-sample R-squared value is obtained by comparing the prediction to the historical mean, as in Campbell and Thompson (2008).

The portfolio returns and PLS allows for significantly positive OOS-R-squared values, outperforming various benchmarks. Once it is established that the composite index of portfolio returns has predictive power, it can be seen as a state variable/risk factor. Then the weights can be examined (which are proven to be stable, diffuse, and reliable while still being different from an equal-weighted approach) to see which type of portfolios contribute more to the state variable. It is found that small-size portfolios and value (high book-to-market) portfolios carry a larger weight on average. The risk factor interpretation is then that low returns on small/value stocks means market returns will be low in the next month as well. This requires a positive premium for investors to invest in these stocks.

Finally, two applications of the newly formed composite portfolio index are performed, which provide a further test of the portfolio weights and predictive information of the forecasts. The first is simply investing in the PLS composite portfolio index, which involves investing using the portfolio weights from each month's forecast. It is found that the portfolio's Sharpe ratio is greater than the market portfolio and an equal-weighted approach. There is much literature (DeMiguel, Garlappi, and Uppal, 2013 for example) that finds an equal-weighted or 1/N strategy outperforms most mean-variance techniques and asset pricing models in terms of out-of-sample Sharpe ratios. This paper provides an endogenous explanation for this phenomenon: the equal-weighted portfolio places more weight on small-size and value stocks compared to the value-weighted portfolio. These stocks carry a different risk premium than the market, which in this case happens to be higher. Going one step further, the PLS composite index places even more weight on small and value stocks, which provides a better Sharpe ratio than the 1/N strategy.

The purpose of the third essay is to examine if imposing the implications of the arbitrage pricing theory (APT) model has out-of-sample benefits. The essay imposes APT pricing restrictions on principal component analysis (PCA). This amounts to forming factors that perfectly explain mean returns (perfect second pass fit) while simultaneously maximizing the covariance of the test assets (maximizing first pass fit subject to the constraint of zero mispricing/alphas). While others have imposed APT restrictions in some way (MacKinlay and Pástor, 2000 and Nardari and Scruggs, 2007), this essay focus on the out-of-sample performance of imposing these restrictions. The essay also differs by using PCA-type factors (termed C-PCA, for constrained PCA), which allows for portfolio weights to be found and applied in real time.

To help evaluate the new approach, a breakdown of mean squared error (MSE) is examined: MSE can be broken down into a variance term and a bias squared term. Obviously, imposing zero mispricing (although the approach allows for a pre-specified allowable percentage of mispricing) will result in a bias of zero in sample. The essay examines the benefits of having an unbiased model, which may be preferable depending on time horizon and investor loss function (asymmetric loss functions should not use MSE as a target). However, the focus is on out-of-sample performance.

Industry portfolios are used as test assets, as they have a weak factor structure, avoiding the Lewellen, Nagel, and Shanken (2010) critique that using portfolios sorted on one or two criteria allows for factors to more easily explain the portfolio returns. While PCA and C-PCA both easily beat the traditional asset pricing models (CAPM, Fama-French three-factor model, and the Carhart model), PCA has the lowest MSE of all models in-sample. The essay shows that this is due to the fact that PCA has an alternative derivation in which the objective is to minimize the total MSE of the panel (in both time and cross-sectional dimensions). C-PCA obviously has the lowest bias in-sample, as that is imposed a priori. However, C-PCA sacrifices only a bit of MSE/variance in order to achieve zero bias.

To judge the out-of-sample performance (ability to estimate expected returns), a hedging approach is implemented. The weights are found from the in-sample estimation window, and are then applied to the next month's realized portfolio returns. This translates to using the factor realizations, and the in-sample betas are then applied to these factor realizations to hedge. If the model held perfectly, then all risk could be eliminated. With an expanding window and an initial window of 50 years, C-PCA again has the lowest bias of all models. C-PCA reduces the bias of PCA by roughly one-third while having only a slightly higher variance and MSE. When the initial window is changed to only 60 months, again bias is reduced by roughly one-third, but the MSE of C-PCA is slightly lower than PCA. A rolling 60-month window is also implemented, but in that case C-PCA does not provide a bias reduction (although it does have the lowest MSE). This is consistent with previous literature (MacKinlay and Pástor, 2000 for example): a longer sample is needed for unbiased estimators to become beneficial.

The main contribution of this dissertation is to show that returns are indeed predictable in some form. It is also shown that this predictability allows for profitable or beneficial applications to be performed. The first essay shows that investor sentiment can be measured in a way that captures only irrationality or overreaction. This applies to the market as a whole, so that market timing provides abnormal positive returns, both statistically and economically significant. The success is due to temporary mispricing resulting from sentiment, followed by market correction in the following month. The second essay shows that a composite index formed by size and book-to-market sorted portfolios can provide significant forecast power. Thus, the index is interpreted as an ICAPM state variable. This offers a potential explanation of the size and value premiums: they reflect future investment opportunities and are therefore riskier. The third paper shows that imposing zero mispricing on asset pricing factors allows for bias to be greatly reduced out of sample. The benefits of doing so are examined, and a potential application of the model and factors is shown. The other contribution of the third essay is to derive this new approach and provide new asset pricing factors for which the APT holds perfectly.

The rest of the dissertation proceeds as follows. Chapter Two contains the first essay ("Forecasting Returns with Fundamentals-Removed Investor Sentiment"), Chapter Three contains the second essay ("Equity Premium Predictions and Optimal Hedging with Many Predictors Using Partial Least Squares"), and Chapter Four contains the third and final essay ("Arbitrage Pricing Restrictions and the Predictability of Stock Returns by Statistical Factor Analysis"). Chapter Five concludes.

References

Baker, M., and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance, 61*(4), 1645-1680.

Campbell, J.Y., and Thompson, S.B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Rev. Financ. Stud., 21*(4), 1509–1531.

Cochrane, J. H. (2005). *Asset Pricing*, Princeton University Press.

de Jong, S. (1993). SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems, 18*(3), 251-263.

DeMiguel, V., Garlappi, L., and Uppal, R. (2007). Optimal Versus Naive Diversification: How Inefficient is the 1/ N Portfolio Strategy? *Rev. Financ. Stud., 22*(5), 1915-1953.

Ferreira, M. A., and Santa-Clara, P. (2011). Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics, 100*(3), 514-537.

Kelly, B., and Pruitt, S. (2013). Market Expectations in the Cross-Section of Present Values. *The Journal of Finance, 68*(5), 1721-1756.

Lemmon, M., and Portniaguina, E. (2006). Consumer confidence and asset prices: Some empirical evidence. *Rev. Finan. Stud., 19*(4), 1499-1529.

Lewellen, J., Nagel, S., and Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics, 96*(2), 175-194.

Li, Y., Ng, D. T., and Swaminathan, B. (2013). Predicting market returns using aggregate implied cost of capital. *Journal of Financial Economics, 110*(2), 419-436.

MacKinlay, C.A., and Pástor, Ľ. (2000). Asset Pricing Models: Implications for Expected Returns and Portfolio Selection. *Rev. Financ. Stud., 13*(4), 883-916.

Merton, R. C. (1973). An Intertemporal Capital Asset Pricing Model. *Econometrica, 41*(5), 867-887.

Nardari, F., and Scruggs, J. T. (2007). Bayesian Analysis of Linear Factor Models with Latent Factors, Multivariate Stochastic Volatility, and APT Pricing Restrictions. *Journal of Financial and Quantitative Analysis, 42*(4), 857-891.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2009). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *Rev. Financ. Stud.*, *23*(2), 821-862.

Chapter Two: *Forecasting Returns with Fundamentals-Removed Investor Sentiment*

Stivers, A. (2015). Forecasting Returns with Fundamentals-Removed Investor Sentiment. *International Journal of Financial Studies*, *3*(3), 319-341.

2.1 Introduction

A common goal in studying investor sentiment is to find whether sentiment can explain returns. Or, do irrational investors cause mispricing that predictably reverses itself in the next period? The key to confirming this hypothesis is finding the right measure of irrational behavior, or investor sentiment. There are many different measures of investor sentiment in the literature, the most popular recently being Baker and Wurgler's (2006) index. As Baker and Wurgler state, mispricing can occur when there are sufficient limits to arbitrage, and uninformed demand shocks occur. Therefore, the ideal sentiment measure would consist entirely of irrational (non-fundamental) investor behavior. Then, this pure measure of investor sentiment should have significant predictive power for future returns. This paper argues that the existing proxies of sentiment fall short in this regard.

The sentiment index used in Baker and Wurgler (2006) may be the most complete measure in the current literature. They create a composite measure of investor sentiment using principal-components analysis. They use six components: closed-end fund discount (from Lee, Shleifer, and Thaler, 1991, but following the methodology of Neal and Wheatley, 1998), market turnover, number of IPOs (Lowry, 2003), average first-day return of IPOs (Derrien, 2005), gross equity share, and a dividend premium measure. However, if one wants a purer measure of sentiment that better captures mispricing, then their sentiment index should be altered. They create a principal-components index based on existing sentiment proxies in the literature, but there are likely to be some rational or fundamental components involved. In order to arrive at a better sentiment measure, the non-fundamental component should be separated from the fundamental components of the Baker and Wurgler index (BW hereafter). However, properly removing as much of the fundamental component as possible is tricky. Baker and Wurgler do attempt to do this by removing what they call business cycle variation from their sentiment measure. They regress each individual component against a dummy variable for NBER recessions, industrial production index growth, and growth in consumer durables, nondurables, and services, and then run principal-components on these six residuals. The sentiment index is still likely to contain rational reactions to fundamentals that do not move with the business cycle. This paper provides empirical confirmation that indeed the BW procedure may not fully remove fundamentals.

A measure of investor sentiment that fully removes fundamentals should do a superior job of forecasting returns due to the following: prices move away from fundamentals when noise trading is correlated and arbitrage is limited, and then deviations are eventually reversed. The basic hypothesis is that mispricing due to sentiment shows up in the aggregate so that market returns are predictable by a pure irrational sentiment measure.[1] This assumes that sentiment drives predictability of returns more than fundamentals, which admittedly may not have a strong prior. However, the evidence in this paper supports the idea of return predictability being due to sentiment and not fundamentals. If sentiment predicts returns only because it captures fundamental factors, then the new sentiment index created in this paper should perform worse in forecasting. However, the opposite is true. Thus, evidence in contrast with some of the literature in this area is provided. It should be noted, though, that Campbell and Kyle (1993) find that mispricing can affect aggregate returns, and Lemmon and Portniaguina (2006) mention that rational and behavioral hypotheses are not mutually exclusive.

It is also important here to note the difference between in-sample explanatory power and out-of-sample predictability. It is common for a model to perform well in sample but poor out of sample. As shown in Section 2.4, the BW index is more significant in sample than the Baker and Wurgler (2006) index that removes business cycle variation. When moving out of sample, though, the BW index performs poorly. The new sentiment index is not intended to fully explain market returns. Rather, the intent is to capture deviations in expected return which are due to irrational mispricing (sentiment). An index that successfully does this should work better out of sample but not necessarily in sample.

This paper removes fundamentals from BW in a way similar to Lemmon and Portniaguina (2006). Lemmon and Portniaguina use a set of fundamentals

[1] This effect is also found in small-size stocks, which is investigated in Section 2.4.4.

that is sufficiently different (and arguably more complete) from Baker and Wurgler's (2006) business cycle variables. Lemmon and Portniaguina separate the fundamental and non-fundamental components of consumer confidence (CC hereafter). They use the non-fundamental component as a proxy for investor sentiment. This paper follows their methodology, but applies it to the Baker and Wurgler index by first regressing CC on a set of fundamentals, then regressing BW on the fitted (fundamental/rational) value from this initial regression, and finally using the residuals from this regression as a proxy for pure sentiment.

Basically, the Lemmon and Portniaguina (2006) measure of fundamental consumer confidence is removed from the Baker and Wurgler (2006) index. Thus, a composite index of fundamentals based on consumer confidence is removed from Baker and Wurgler's composite index of sentiment. This composite index of fundamentals should limit any measurement errors or over-fitting of fundamentals when removing them from BW.[2] Also, Lemmon and Portniaguina's set of fundamentals (which this paper uses with some minor changes) is a more complete measure of overall fundamentals compared to Baker and Wurgler's business cycle variables. This paper shows that, empirically, following this methodology allows investor sentiment to forecast market returns better.

Instead of directly following Lemmon and Portniaguina (2006), this paper uses BW for forecasting instead of CC, as it is a market measure of investor

[2] The large number of fundamental variables and their lags may result in over-fitting, which is why the principal-components method is commonly used with a large number of variables. PC is also used to forecast, with inferior results shown in the Appendix.

sentiment. [3] Market measures of sentiment are composed of real actions by investors directly involved in the market. For surveys (especially consumer confidence surveys), a participant may feel overly optimistic but may not actually act on this optimism. Surveys are also open to problems such as dishonesty, bias, and incomplete responses. A market measure of sentiment (such as BW) will match investors' actions as it relates to their overall feeling of the stock market. Thus, this paper proposes that a market measure of investor sentiment will forecast market returns better than a survey-based measure. Empirical evidence is presented in support of this hypothesis, as the new index performs better in forecasting market returns than both Baker and Wurgler's (2006) index and their alternative sentiment index that removes business cycle variations (BWA hereafter).

To measure the predictive power of the new index, a linear forecasting model is used to forecast one-month-ahead market returns. The sample runs from July 1978, when the University of Michigan consumer sentiment survey became available in a monthly frequency, to December 2010. A simple investment strategy is used: if the one-month-ahead predicted excess market return is higher than the historical average of all previous months in the sample at the time of forecasting, then the investor will invest in the market. Otherwise, she invests in a risk-free asset (T-bills). Essentially, the goal is to time the market. This differs from both Baker and Wurgler (2006) and Lemmon and Portniaguina (2006), where no market timing is involved.

[3] It is also shown in this paper that Lemmon and Portniaguina's (2006) measure of sentiment does not predict returns as well as the new, modified BW index.

Market timing is considered here for two reasons. First, it is a practical application of investor sentiment, in that it is usable by practitioners. Since a common goal in the sentiment literature is predicting returns, market timing is a natural extension. The second reason for using a market timing approach is that it provides economic significance (via realized returns) in addition to statistical significance. Following the investment strategy (essentially switching between stocks and bonds), the forecast model that removes the fundamental component of CC from BW provides average realized excess returns of around 10.6% annually compared to around 5.3% annually for BW and 8.0% annually for BWA. This new sentiment index also has significant market timing ability, as will be shown later. The excess market return average over the forecasting sample is around 6.7% annually (the return from a buy-and-hold strategy), so a forecast model utilizing only BW would not beat the market on average. This new index performs significantly better than both CC and Lemmon and Portniaguina's sentiment component of CC (by itself) as well. Since sentiment is typically thought to affect small, young, and volatile stocks the most, forecasting is also performed on a small-size portfolio. Again, removing the fundamental component increases the realized returns an investor could obtain. The evidence supports the hypothesis that investor sentiment should have no fundamental (rational) component, and that the existing measures of sentiment do not properly remove fundamentals. It should be noted that predictability in returns may be driven somewhat by time-varying risk and risk aversion, and not entirely by sentiment. However, this paper provides evidence that creating a better measure of irrationality increases predictability in returns.

The paper proceeds as follows: Section 2.2 reviews the existing literature on investor sentiment and consumer confidence as a proxy for sentiment; Section 2.3 discusses the data and methodology used; Section 2.4 discusses the results; and Section 2.5 concludes the paper.

2.2. Literature Review

DeLong *et al.*, (1990) show that returns of assets mostly held by noise traders can be predictable as mispricing caused by correlated sentiment (with limited arbitrage) will eventually correct itself. [4] Most studies (Baker and Wurgler, 2006, Lemmon and Portniaguina (2006), *etc.*) show that small and young firms (and closed-end funds, discussed shortly) are predominantly held by noise traders and thus are more susceptible to sentiment. However, Campbell and Kyle (1993) show that overreaction can impact aggregate stock values as well. Hence sentiment may also be able to predict market returns.

Lee, Shleifer, and Thaler (1991) show that closed-end fund discounts may be due to investor sentiment. In their study, the returns of ten size-ranked portfolios are regressed against market returns and the change in a value-weighted closed-end fund discount variable. They find that the smallest size-ranked portfolio moves closely with closed-end funds, while the largest size-ranked portfolio moves in the opposite direction of closed-end funds.

[4] For a survey of behavioral finance, see Hirshleifer (2001).

The papers that follow Lee, Shleifer, and Thaler (1991) are at the core of the field of investor sentiment. Barberis, Shleifer, and Vishny (1998) present a model of investor sentiment that supports the findings that stock prices underreact to earnings announcements and overreact to successive good or bad news. Elton, Gruber, and Busse (1998) show that changes in closed-end fund discounts do not explain common stock returns when including a value-weighted industry return index. They also find that investor sentiment is not a priced factor in common stocks as well as closed-end funds. Neal and Wheatley (1998) find that closed-end fund discounts can predict the size premium, but nothing else.

Other authors have considered proxies or measures of investor sentiment outside of closed-end fund discounts. Lee, Jiang, and Indro (2002) use a GARCH-mean model and the Investor's Intelligence survey as their sentiment proxy to show that sentiment is a priced systematic risk. Lowry (2003) shows that investor sentiment may impact IPO volume. Cai *et al*. (2013), find that sentiment affects variation in straight debt IPO volume, and Derrien (2005) shows that sentiment may play a role in the initial return of IPOs. Baker and Stein (2004) create a model in which high market liquidity indicates that the market is dominated by irrational investors, creating a role for sentiment.

There is a wide range of literature in which consumer confidence indices are used to proxy for investor sentiment. Fisher and Statman (2003) find that consumer confidence impacts individual investors' sentiment, but not institutional investors' sentiment. They also find that consumer confidence will rise significantly when stock returns (using many different indices) are high, and that consumer confidence may have some predictive power for returns. Schmeling (2009) uses consumer confidence to try to predict returns for 18 different countries. He finds that consumer confidence has predictive power up to 6 months, but dies out after that. Also, some countries' consumer confidence measure has no predictive power whatsoever.

Lemmon and Portniaguina (2006) use consumer confidence as a proxy for investor sentiment. As mentioned earlier, Lemmon and Portniaguina (2006) create a sentiment proxy from consumer confidence by removing nine quarterly fundamentals and their lags. They point out that the rational hypothesis and sentiment hypothesis are not mutually exclusive. Therefore, by removing fundamentals they create a sentiment index similar to what this paper creates: an index where mispricing could only be due to sentiment, so that prices will reverse and be predictable. They use a quarterly index of the University of Michigan Consumer Sentiment Index (CC). They mention that from 1978–2002 their sentiment component and the Baker and Wurgler (2006) index have a very small correlation and that their index does not support Baker and Wurgler's findings mentioned below. They do not attempt to predict market returns as this paper does. The other key difference in this paper is that here their fundamental component of CC (using monthly data) is then removed from BW (a more direct market measure) to be used in market timing.

As discussed in the previous section, Baker and Wurgler (2006) create a composite investor sentiment index using principal-component analysis. They use their index to run a predictive regression on various long-short portfolios.

Their main findings are that small, young, and volatile firms have low returns for the following year when current sentiment is high. Baker and Wurgler (2007) create a first-differenced series of their 2006 sentiment indices, whereby they first-difference the individual components and then run principal-components analysis. They use the change index to test for return co-movement with sentiment, while using their 2006 levels index to forecast returns. Their findings are that speculative, difficult-to-arbitrage stocks and stocks with high volatility are impacted greatly by sentiment. They also find that high sentiment leads to lower future market returns (overreaction). This paper builds on this finding, and attempts to forecast market returns.

2.3. Data and Methodology

The sample period used in this paper runs from July 1978 (when the Michigan Survey began to consistently use monthly frequencies) to December 2010, using monthly data. Market return is the value-weighted CRSP measure that includes dividends. [5] The risk-free rate used is the rate on 3-month Treasury bills, obtained from the St. Louis Federal Reserve (FRED) website.[6] Excess market return is the difference between these two. The University of Michigan Consumer Sentiment Index (CC, or consumer confidence) is also obtained from the FRED website. The Baker and Wurgler (2006) index and its alternative that removes business cycle variation, along with the individual

[5] Other market indices were examined as well with qualitatively similar results. The S&P 500 index was used, as well as the value-weighted NYSE/AMEX measure that excludes REITs, closed-end funds, and ADRs from Statman, Thorley, and Vorkink (2006).
[6] Available online: https://research.stlouisfed.org/

components, are obtained from Jeffrey Wurgler's website.[7] Table 2.1 shows the descriptive statistics for these variables. Figure 2.1 provides time-series graphs of the excess market return, CC, BW, and BWA over the sample for comparison purposes.

[Table 2.1 goes here]

As can be seen from Table 2.1, the average market return for the sample used here is about 12.3% annually (1.027% monthly). The average excess market return is slightly less than 7% annually. CC uses 1966 as its base year, with CC = 100. BW is from a standardized principal-component analysis, so that the series (although not necessarily in the sample used here) has a unit variance and a mean of zero. As described in the literature review, BW is comprised of six individual sentiment components and employs both leads and lags.

While CC is definitively non-stationary, BW's stationarity is ambiguous using Dickey-Fuller, Dickey-Fuller GLS, and Phillips-Perron tests. For BW, the Dickey-Fuller GLS test shows that it is non-stationary, but the other tests are inconclusive. It appears that BW may be non-stationary due to the non-stationarity of just one component, the turnover component. Further, CC and BW may be cointegrated, using the Johansen integration test. The results are not definitive, as would be expected since BW may or may not be non-stationary.

[Figure 2.1 goes here]

[7] Available online: http://people.stern.nyu.edu/jwurgler/

Initially, Lemmon and Portniaguina's (2006) methodology is implemented, with some minor changes. Primarily, Lemmon and Portniaguina use quarterly data and this paper uses monthly data. Lemmon and Portniaguina use the following nine fundamentals: default spread (DEF), as measured by the difference between the yields to maturity on Moody's Baa-rated bonds and Aaa-rated bonds; yield on 3-month Treasury bills (RF); dividend yield (DIV), following Fama and French (1988); real GDP growth (GDP); growth in personal consumption expenditures (CONS); labor income growth (LABOR), as measured by the per capita growth in total personal income minus dividend income, deflated by the PCE deflator; the Bureau of Labor Statistics unemployment rate (URATE), seasonally adjusted; inflation rate (INF), measured by the change in the consumer price index; and the consumption-to-wealth ratio (CAY) from Lettau and Ludvigson (2001). The dividend yield measure can be obtained from the difference between CRSP's value-weighted return index and its value-weighted return index excluding dividends. The CAY measure is obtained from Martin Lettau's website.[8] The seven other fundamental variables can be obtained from FRED's website.

This paper makes two changes to Lemmon and Portniaguina's (2006) choice of fundamentals: instead of using the unemployment rate the prime rate on bank loans (PRIME) is used, and CAY and GDP are interpolated to monthly data. PRIME is obtained from FRED. The prime rate is essentially a benchmark for other loan rates, so it should broadly capture fundamental economic activity. Also, the unemployment rate is a survey measure, while the

[8] Available online: http://faculty.haas.berkeley.edu/lettau/data.html

other fundamentals (and the prime rate) are directly observable. [9] The prime rate is more quickly accessible to investors than is the unemployment rate as well. Following Lemmon and Portniaguina, CC is regressed against these 9 fundamentals and their lags using monthly data. Since Lemmon and Portniaguina use quarterly indices with one lead and one lag (giving six months of data), the lead on each fundamental along with 5 monthly lags is used. Since CAY and GDP are converted to monthly, only a lead and one 3-month lag can be used for these two variables to avoid multicollinearity. This gives the following:

$$CC_t = \alpha_0 + \sum_{i=0}^{5}[\alpha_i]*[X_{t-i}] + v_t = CCfit_t + v_t \qquad (2.1)$$

α_0 is a constant, $[\alpha_i]$ is a vector of coefficients:

$[X] = [DEF, RF, DIV, CONS, INF, LABOR, PRIME, CAY, GDP]', CCfit_t$ is the fitted value of the regression, and v_t is the error term (noting that CAY and GDP only have one lag at t - 3).

The fitted value from the above equation is obtained by multiplying the coefficient by the respective value for that variable (from i = 0 to 5) and summing them together with the intercept for each particular month. This produces the "fundamental" or "rational" component of CC, whereas the residual is the sentiment component as in Lemmon and Portniaguina (2006). As discussed in the introduction, the idea is to weight the fundamentals on the

[9] In addition to possible survey errors and biases, the debate over how an unemployment statistic should be constructed leads to potential problems in its use.

non-market survey measure of consumer confidence before removing it from BW. The aim is to remove the rational, fundamental component of BW, leaving only irrational investor sentiment (which produces market mispricing).[10]

Each regression is rolled forward one month, while anchoring the starting point. This is done so that only information from time 1 to t is utilized in forecasting at month $t + 1$. Therefore, the coefficients will change as the regression updates forward each month in a stepwise fashion. The fitted values of Equation (2.1) are saved each time so that a series is created to be used in forecasting. That is, Equation (2.1) is run from time 1 to time t, with t fitted values created. This series can be used to then forecast at $t + 1$, as will be explained later. To forecast the next month ($t + 2$), Equation (2.1) is then run from time 1 to $t + 1$, creating $t + 1$ values in each series. Thus, Equation (2.1) is updated each month as forecasting moves from the halfway point to the last period in the sample. When running Equation (2.1) for the entire sample, the adjusted R-squared value is 0.78, which is very close to the value that Lemmon and Portniaguina (2006) obtain.

The fitted (fundamental) CC values are then used to properly remove fundamentals from BW. Again, BW is used to forecast as it is a market-based measure and should forecast market returns better. Lemmon and Portniaguina's methodology has been used up to this point in order to get a

[10] Removing the fundamental component of BW from CC does not work well when forecasting. This means switching CC and BW in Equations (2.1) and (2.2). Results are available upon request.

proper measure of fundamentals. Now, after obtaining the series of fitted values from Equation (2.1), the following is run:

$$BW_t = a + b * CCfit_t + e_t \tag{2.1}$$

$$e_t \equiv FRS_t \tag{2.1}$$

CC*fit* is the fitted value of CC from Equation (2.1). The procedure in Equation (2.2) is the same as in Equation (2.1). The residual from Equation (2.2), e_t, is hereafter referred to as FRS (fundamental-removed sentiment). Note that using only CC*fit* eliminates any endogeneity issues that would arise from using CC in Equation (2.2). The reason is that since CC also has a sentiment component, it would be correlated with the residual from Equation (2.2), or FRS. [11] CC*fit* is used as a composite index of fundamentals which should help prevent potential over-fitting or measurement errors when removing fundamentals from sentiment. Table 2.2 shows the descriptive statistics for FRS, when forecasting starts at the one-third mark of the sample as will be done later. Note that this is the series of only time *t* residual from Equation (2.2), where time *t* is used to forecast excess market return at time *t* + 1. Since FRS contains residuals from different regressions, it is not necessarily expected that its mean be zero (as opposed to a series of residuals from one regression). It should also be noted that FRS is unambiguously stationary when BW is modified using this methodology. As discussed at the beginning of this section, BW is found to be non-stationary using some statistical tests.

---

[11] One other possible benefit from using CC*fit* as a composite index of fundamentals is that directly removing fundamentals instead may remove both a rational and irrational component from BW (the irrational part being an overreaction to a fundamental shock).

[Table 2.2 goes here]

The regression in Equation (2.2) is run in a stepwise fashion with an anchored starting point, with the end point starting from the halfway point (October 1994) updating each time to include the next month. Equation (2.2) can be thought of as running simultaneously with Equation (2.1). The residuals from Equation (2.2) are then used in forecasting excess market returns, with each month in which forecasting is done having its own unique series of the residuals from Equation (2.2) that does not include any future information. The R-squared value from Equation (2.2) is 8% for the full sample, although it ranges from 3%–8% over the sample.

The following forecasting models are employed:

$$\text{Model i: } \hat{R}_t = \alpha_t + \beta_t * sent_{t-1}^{i} + \eta_t \quad \text{i = 1,2,3} \qquad (2.2)$$

$\hat{R}_t$ is the predicted excess market return and $sent_{t-1}^{i}$ is the appropriate sentiment measure, lagged one month, for model i. For Model I (i = 1) BW is used for “sent”, for Model II (i = 2) BWA is used, and for Model III (i = 3) FRS is used. BWA is included for comparison purposes; this also shows that BW removes some fundamentals but ultimately falls short. Forecasting is done from October 1994 to December 2010 and also from May 1989 to December 2010. Therefore, out-of-sample forecasts for the second half of the sample are obtained, and also for the last two-thirds of the sample for comparison and robustness. So when forecasting starts at the halfway point of the sample, the forecasting equations above start by running from July 1978–

September 1994, and finish by running from July 1978–November 2010. Ordinary least squares is used in the forecasting equations, and one-month-ahead forecasts of the excess market returns, $\hat{R}_t$, are obtained using the residuals of the forecasting equations.

2.4. Results

2.4.1 Forecast Results

Table 2.3 shows the results from the forecasting regressions for each of the three models (for the first case when forecasting starts at the halfway point of the sample). The coefficients, *t*-statistics (reported in parentheses), and R-squared values in the table are averages from the 195 individual forecasting regressions for each model, with each regression adding one additional data point. T-statistics are obtained by following the Fama-MacBeth (1973) methodology, using the distribution of the 195 coefficients. Note that these *t*-statistics should not be biased since the standard two-pass approach with generated regressors is not used here. Excess market returns are the dependent variables in each model and its respective regressions. Coefficients for each variable are small, but the coefficients are difficult to interpret in terms of size due to how CC and BW are created, as well as the fact that residuals from Equation (2.2) are used as independent variables.

In-sample R-squared values are low for each model, as is common in return forecasting. Campbell and Thompson (2008) provide a way to examine whether a model provides sufficient explanatory power. Using out-of-sample

(OOS) R-squared values based on the historical mean, the OOS R-squared value can be compared to the squared Sharpe ratio (here of the excess market return) over the forecasting sample. When the time interval is small (monthly is small enough) then a mean-variance investor can use Model III (FRS) to increase her monthly expected portfolio return by a proportional factor of R^2/SR^2 (out-of-sample R-squared over the squared Sharpe ratio for the forecasting sample). Here, Model III gives 1.3/1.6 = 82% proportional increase in expected return from observing FRS.

Model III, when removing CC*fit* from BW and using this residual to forecast excess market returns (FRS), is the most significant, with the most negative coefficient. FRS is a measure of uninformed demand shocks, which cause mispricing and subsequent reversal. A one standard deviation increase in FRS for this particular subsample implies a 0.25% decrease in next month's excess market return. Thus, sentiment and excess market returns have a negative relationship. This can be interpreted as overreaction on the part of investors.

[Table 2.3 goes here]

Table 2.4 shows the forecast results. To measure forecast (market timing) accuracy, a simple investment strategy is implemented. The one-month-ahead forecast of excess market return is compared to the historical average of the excess market return. So if the predicted value of the excess market return ($\hat{R}$) is calculated at time t + 1, the relevant historical average of the excess market return is calculated from time 1 to t. If $\hat{R}_{t+1}$ is higher than the average obtained from existing data at that point (at time *t*), then the investor will invest only in the market in the following month ($t + 1$), obtaining the actual market return in that period. However, if $\hat{R}_{t+1}$ is less than the historical average, then the investor will invest only in the risk-free asset in the following month, obtaining the actual risk-free return. This produces a time series of realized investment returns for each forecasting model. That is what this paper is concerned with; it produces a measure of the ability of each model to time the market. This is also the return an investor could conceivably have earned when using the respective model to forecast based on the investment rule, since out-of-sample forecasting is done. Panel A of Table 2.4 provides for each model the realized investment excess return, standard deviation, and risk-corrected average realized investment returns. Note that realized investment returns are either market returns or risk-free returns (minus the risk-free rate). Returns reported in Table 2.4 are annualized monthly returns.

[Table 2.4 goes here]

From the table it can be seen that using the Baker-Wurgler (2006) index to forecast (Model I) provides average realized investment excess returns of about 5.3% annually (8.6% prior to subtracting the risk-free rate), and a realized standard deviation of about 3.5% (annualized Sharpe ratio of 0.44). The alternative Baker and Wurgler index that removes business cycle variation (Model II) provides average excess returns of 8.0% (11.3%) and a standard deviation of 3.2% (annualized SR = 0.72). However, one can do even better when removing the fundamental component of CC from BW (Model III, using FRS), with an average annual excess return of about 10.6% (13.9%) and a standard deviation of 3.8% (SR = 0.81). Model I does not beat the average market excess return for the forecasting sample period, which is around 6.7% annually. Note that this 6.7% would be the realized excess return from a buy-and-hold the market strategy, and the strategy would have a standard deviation of 4.8% (SR = 0.40). Model II beats a buy-and-hold strategy by 1.3%, and Model III beats a buy-and-hold by 3.9% for this sample. As mentioned previously, it is possible that returns are predictable in part due to time-varying risk and risk aversion and not solely due to investor sentiment, though performance increases as more fundamentals are removed. The results are even more divergent when using the median rather than the mean of realized returns, which is shown in the second row of Panel A. Models I and II drop to a 1.8% realized return, while Model III has a median of 9.4%. Models I and II produce realized returns that are skewed (more negative returns compared to Model III). In order to check if the positive results of Model III are due to a few outstanding months' returns, the top 1% is dropped and the average of the remaining 99% is reported in the third row of Panel A. While the returns in row 3 drop across the board (as to be expected), they still remain significantly positive at 1% for Model III.

Pesaran and Timmermann (1994) raise the issue that a market timing portfolio that switches between stocks and bonds may have a lower standard deviation simply due to including bonds with low standard deviation. To address this, they construct a portfolio that randomly switches between stocks and bonds at the same proportion as their switching portfolio. The expected excess return of this portfolio with the CRSP market data and T-bills used in this paper would be 5.2% with a standard deviation of about 5.0%. Note that Model III provides a superior mean-variance tradeoff.

Also, BW does even worse in comparison after correcting for market risk, with an average annual return of about 1.7%. FRS gives market-risk-corrected returns of about 6.4% annually. BWA gives market-risk-corrected returns of about 5.0%, again falling below Model III but well above Model I (BW). The results stand after correcting for Fama and French's (1996) HML and SMB factors, and Carhart's (1997) momentum factor as well. The size factor has been suggested as possibly explaining abnormal returns from a sentiment-based trading strategy, but these results show otherwise. Model III (FRS) still provides an average annual return of 5.8% after correcting for market, size, and value. Momentum also does not fully explain the performance of FRS.

The last two columns of the table show the difference in returns between Model III and the other two models (or the improvement of FRS over BW and BWA). Significance levels are also shown, which are obtained from a bootstrapping procedure with 10,000 simulations. FRS offers an improved return at a 5% significance level over BW and at a 10% significance level over BWA. This is the case for both average returns and returns with the top 1% removed. For median returns, FRS offers an improvement at a 1% significance level over both BW and BWA. FRS offers a significant improvement over BW when using risk-adjusted returns. While the risk-adjusted returns of FRS are not a significant improvement over BWA using the bootstrapping procedure, there are two important points to consider. The first is that these returns are less volatile with a smaller range for all three models. Therefore, significance is more difficult to achieve. The second point is that the difference in risk-corrected returns between Models III and II are still economically substantial, with the difference in being between 1.2% and 1.4% per annum (note that this is essentially the alpha obtained).

2.4.2. Market Timing Tests

Looking at the realized returns that each model would provide for an investor provides a test of economic significance, but the market timing literature provides additional statistical tests for market timing ability. Three such tests are performed here for Model III over the October 1994–December 2010 forecasting sample, and the results can be seen in Panel B of Table 2.4. The first is the Treynor and Mazuy (1966) test (TM hereafter), where a quadratic specification is implemented. The TM test for market timing ability is testing the null hypothesis that the coefficient on the squared excess market return is not positive, when the realized excess returns are regressed on a constant, the excess market return, and the squared excess market return. From Panel B it can be seen that Model III (FRS) has a large positive coefficient on the squared excess market return, and it is significant at 10% based on Newey-West (1987) standard errors. Thus, there is significant market timing ability based on the TM test. [12]

Henriksson and Merton (1981) argue that market timers have different target betas depending on whether they predict an up or down market,

[12] For all three market timing tests, the market timing ability coefficients are all significant at 1% when the Newey-West (1987) correction is not performed.

therefore their test (the HM test) replaces the squared excess market return variable of the TM test with the following dummy variable:

$$D_t = \max[0, (rf_t - R_t)] \tag{2.3}$$

rf is the return on 3-month U.S. Treasury bills, and *R* is the excess market return. Panel B shows that the coefficient (timing ability) on this dummy variable is positive and significant at 5%. While the intercept in this specification can be interpreted as selection ability, Goetzman, Ingersoll, and Ivković (2000) note that the selection ability coefficients and market timing ability coefficients are typically of the opposite sign.

The Cumby and Modest (1987) test (CM), but following Breen, Glosten, and Jagannathan (1989) is also performed. They do not assume a CAPM structure and directly test whether the expected realized return is different when an up market is predicted compared to when a down market is predicted. The dummy variable for this test (the variable I in Panel B of Table 2.4) is equal to one if the predicted return of Model III is higher than average, and equal to zero if it is less than average. Here, the excess market return is the dependent variable. Again, the market timing ability coefficient is positive and significant at 1%, providing positive evidence of timing ability. Breen, Glosten, and Jagannathan (1989) also note that a negative intercept is equivalent to the expected return of the timing portfolio being higher than the return on the market (buy-and-hold strategy). Panel B shows a negative intercept that is significant at 10%.

In order to truly examine the potential profitability of the model, transaction costs must be taken into account. However, the emergence of ETFs has allowed for low transaction costs when switching between a market proxy portfolio (ETF) and a bond (possibly an ETF as well). For example, Vanguard advertises a 0.15% average ETF expense ratio, although it notes that the industry average according to Lipper, Inc. is 0.58% as of December 31, 2012. Vanguard's CRSP total stock market index has an expense ratio of only 0.05%, but other mutual funds may be as high as 0.89%.[13] Table 2.5 shows the realized returns of each model after accounting for transaction costs. The high and medium numbers come from Pesaran and Timmermann (1994). They use a cost of 1% and 0.5% for the high and medium, respectively, transaction costs of stocks. These are close to the industry average of 0.58% and the upper end of mutual funds of 0.89%. Also, a low cost of 0.15% is given (Vanguard's cost). A cost of 0.1% is used for bonds in all cases, as in Pesaran and Timmermann. The costs are applied to stocks and bonds for each month in which the investor would hold the market (stocks) or the risk-free asset (bonds). From examining Table 2.4 it can be seen that the realized returns are not significantly lowered after accounting for transaction costs. [14]

[Table 2.5 goes here]

[13] Obtained from Vanguard's website. Available online: https://investor.vanguard.com/home/

[14] Note that these are average returns and not cumulative. Transaction costs would have a larger effect when examining cumulative returns, as the lower returns are then reinvested each month. See Pesaran and Timmermann (1994) for more details.

Several other forecast models are examined to provide a comparison. The results can be found in the Appendix. [15] None of the models examined come close to the performance of FRS. Interestingly, directly removing the fundamentals in Equation (2.1) from BW does poorly as well. Thus it appears that the CC index properly weights these fundamentals, which when removed from BW provide a better irrational measure of sentiment that also forecasts better.[16]

2.4.3. Statistical Significance

In order to obtain a general idea of how well the investment strategies perform, the results from Model III are compared to random strategies that pick between the market and the risk-free asset randomly each month. A simulation is conducted where the market is picked about 80% of the time, as is the case with Model III. [17] Under the null hypothesis that excess market returns are not predictable (efficient market hypothesis), this would give an accurate distribution. Thus, our null hypothesis is the rational efficient market hypothesis, and our alternative hypothesis is the behavioral hypothesis (mispricing caused by noise trading). This analysis can also be seen as comparing the appropriate forecasting model to a random walk model. When performing the random draws 20,000 times, the percentage of times that the

[15] One of the other variables tested in the Appendix is the Lemmon and Portniaguina (2006) sentiment measure, which gives inferior results.

[16] Forecasting was also performed with both the Lemmon and Portniaguina (2006) measure of sentiment (the residual v from Equation (2.1)) and FRS in the forecast regression. The result is a similar, albeit slightly lower, realized return than when using FRS alone (Model III).

[17] Note that the expected return of a portfolio in this simulation is the same as that from Pesaran and Timmermann's (1994) methodology, described in the previous subsection.

realized investment return beats Model III is indistinguishable from zero. In fact, only 14 out of 20,000 simulations beat Model III. This could be interpreted as a *p*-value of zero under the null hypothesis. Hence, the realized return of Model III would be significant at 1%, as shown in Table 2.4. For this model, the 1% cutoff is 9.0% average annual excess return, and for the 5% level it is 7.9% (compared to Model III's performance of 10.6%). Following the same methodology (while changing the probability that the market is selected to match the appropriate model being used for comparison), Model II is beaten 2.8% of the time (significant at 5%) and Model I is beaten 24% of the time (insignificant). The same is done for risk-corrected returns, and those significance levels are denoted in Table 2.4 as well.

2.4.4. Robustness

Forecasting is also performed starting at the one-third mark of the sample (May 1989). This provides 260 months of forecasting one-month-ahead excess market returns, instead of the 195 produced earlier. Table 2.6 shows that the overall forecast results for the full forecasting sample (which is now two-thirds of the entire sample) are qualitatively the same as in Table 2.4. Again, removing the fundamental component of CC from BW provides the highest realized excess returns (Model III is again superior). In this longer sample period, the performance of BW/Model I can be completely explained by the known risk factors, while the performance of FRS is still significantly positive after risk correction. This is further support that Model III provides superior forecasting performance and realized investment returns. A buy-and-hold the market strategy would have provided a realized excess return of about 6.5% (standard deviation of 3.2%), so again Model I does not beat the market, while Model II is on par with the market. Model III again clearly outperforms the market. Model III again is clearly superior to the other two models when examining the median of realized returns rather than the mean. Also, the results are not due to a few positive outliers, as can be seen from the third row of the table, which drops the top 1% of realized returns and then takes the mean.

[Table 2.6 goes here]

Also, the same significance check from Section 2.4.3 is performed for this sample and is reported in Table 2.5. Model I is outperformed by random strategies (with the same probability of being in the market) 29% of the time, Model II 9.25% of the time, and Model III approximately 0% of the time (only 5 out of 20,000 simulations were superior). The significance levels for Models I and III are unchanged with this longer forecasting sample, and now Model II is only significant at 10%. So out-of-sample market timing for this larger window is again clearly superior when removing fundamentals from BW. Further, it is clear when examining the results that forecasting (market timing) improves as time goes on, perhaps due to increased data to be used in forecasting out of sample. For example, from 1989–1994 Model III provided about the same return as the market. From 1995–1999, Model III slightly underperformed the market, and from 2000–2005 Model III beat the market by 4.5% annually. However, from 2006–2010 Model III beat the market by 9.3% annually. For this time period, Model III also provides a market-risk-corrected return of about 10.5% annually. The last few years are not the sole driving force for full sample performance, however. From 1994–2002 Model III still beats the market by 2.3% annually, while Model II (BWA) underperforms the market by about 0.7% annually in this subsample. Again, the last two columns show Model III's (FRS) improvement over BW and BWA. While most of the differences are qualitatively the same as Table 2.4, FRS is now comparatively stronger than BWA. FRS now offers average returns that are higher than BWA at a 5% (rather than 10% as before) significance level. Also, the Fama-French (1996) 3-factor risk-corrected returns of FRS are now significantly higher than BWA.

One possible explanation for the increasing return may be that the time-varying systematic risk is increasing as well. While it is true that the standard deviation of the market increases for each subsample, the Sharpe ratios from the investment strategy based on Model III increase relative to the market over time. For the 1989–1994 subsample, the investor would have obtained about the same Sharpe ratio as the market. The subsample with the highest SR is 1995–1999 (1.32) but it is slightly below the market SR of 1.41. For 2000–2005 the investor would have obtained a realized SR of 0.29 compared to the market's negative SR, and for 2006–2010 the SR is 0.89 compared to the market SR of 0.17. The 2006–2010 subsample survives market risk correction as well, as mentioned in the previous paragraph. Based on the Sharpe ratios, it looks like the 2000–2005 subsample performs just as well as the 2006–2010 subsample compared to the market (or a buy-and-hold strategy). Also, the 1995–1999 subsample provides a slightly higher standard deviation to the investor than the 2006–2010 subsample, but the former does not beat the market while the latter beats the market by over 9% annually. So it would seem that time-varying systematic risk is not the driving factor in the increasing performance.

The *t*-statistics, coefficients, and R-squared values from the forecasting regression all increase in absolute value over time as well. Figure 2.2 shows a 12-month moving average of the R-squared for the forecasting regression of Model III, starting at the halfway point of the sample. Here, the forecasting regression is done with a rolling, fixed-window regression without an anchored starting point, so that 195 months of data are used to show the R-squared value of the regression ending at the appropriate month in the figure. Figure 2.2 shows an upward trend in the R-squared values, which shows that the FRS model does better the later in the sample forecasting is performed. This is an interesting point to examine in further research. What are the reasons for the increasing performance? One possibility is that sentiment (specifically, the arguably purer measure of sentiment, FRS) is playing a larger role in the market. Also, it will be interesting to see how this model performs in the future, if it can keep up the escalating performance.

[Figure 2.2 goes here]

A further robustness check is performed by forecasting a small-size portfolio. Since noise traders (sentiment) tend to hold mostly small, young, and volatile firms, the new sentiment index should also predict small-cap stocks well. The same out-of-sample analysis is performed as was done for market returns: the three models are used to forecast a small-stock portfolio.

Based on the forecast (above or below the trailing historical average), either the small-size portfolio or risk-free asset (respectively) is chosen for the upcoming month. Then the series of realized returns can be analyzed. The smallest size decile from Kenneth French's size-sorted portfolios is used as a proxy for small-cap stocks.[18] The sample is the same as for the market analysis: estimation starts in July 1978 and forecasting is performed from October 1994 to December 2010. One-month-ahead recursive forecasting with an anchored starting point is performed again.

The results can be seen in Table 2.7. Model III (FRS) again outperforms BW and BWA. From examining the average annualized excess returns, Model III outperforms a buy-and-hold the market strategy and a buy-and-hold the size portfolio strategy (which would earn an investor a 10.8% annualized excess return on average). FRS performs better than BW and BWA after correcting for risk as well. The returns for Model III are higher and more significant than Model I and II across the board. The significance levels are again obtained by simulating realized returns with the same average investment proportion. It should be noted that Model III does take on a bit more risk than the other models, and Model III's return becomes closer to Model II's return as more and more risk factors are taken into account.

[Table 2.7 goes here]

[18] Data obtained from Ken French's website. Available online: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

2.5. Conclusions

Removing fundamentals from the Baker and Wurgler (2006) investor sentiment index by first weighting them on the University of Michigan Consumer Sentiment Index (regressing CC on the fundamentals and taking the fitted value) provides better forecasting and higher realized investment returns based on a strategy that utilizes the forecasts. Also, a forecast model that uses the newly constructed sentiment index shows significant market timing ability. The evidence in this paper provides support that the fitted value of CC (regressed on fundamentals) is a proper composite index of fundamentals, and would prevent over-fitting and/or measurement errors when removing fundamentals from sentiment. Support for the behavioral hypothesis of DeLong *et al.*, (1990) is presented in this paper. It appears that removing fundamentals from BW creates a better measure of non-fundamental demand shocks. Under the behavioral hypothesis, these uninformed demand shocks cause market mispricing followed by reversal, allowing for better market timing. Particularly, it better captures correlated noise amongst investors that is not related to any macroeconomic condition (or rational response to these conditions). This provides evidence that sentiment may be a better predictor of returns than fundamentals. The new index also has the added benefit of being conclusively stationary, which cannot be said about the original Baker-Wurgler (2006) index. The new measure predicts a small-cap portfolio better than the two Baker and Wurgler measures as well. The evidence presented in this paper supports the view that the two variants of the widely used Baker and Wurgler sentiment index do not properly remove fundamentals. A fully irrational (non-fundamental) measure of investor behavior is needed to properly test the behavioral hypothesis of sentiment causing mispricing and predictable returns, and this paper attempts to create such a measure. This paper presents evidence in favor of irrationality in the stock market (investor sentiment), and evidence against market efficiency. This paper also extends the work of Lemmon and Portniaguina (2006), by implementing their methodology on the market sentiment measure of Baker and Wurgler (2006), and forecasting monthly returns.

References

Baker, M., and Stein, J. C. (2004). Market liquidity as a sentiment indicator. *Journal of Financial Markets, 7*(3), 271-299.

Baker, M., and Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance, 61*(4), 1645-1680.

Baker, M., and Wurgler, J. (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives, 21*(2), 129-151.

Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics, 49*(3), 307-343.

Breen, W., Glosten, L. R., and Jagannathan, R. (1989). Economic Significance of Predictable Variations in Stock Index Returns. *The Journal of Finance, 44*(5), 1177-1189.

Cai, K. N., Jiang, X., and Lee, H. W. (2013). Debt IPO Waves, Investor Sentiment, Market Conditions, and Issue Quality. *Journal of Financial Research, 36*(4), 435-452.

Campbell, J. Y., and Kyle, A. S. (1993). Smart money, noise trading and stock price behaviour. *The Review of Economic Studies, 60*(1), 1-34.

Campbell, J. Y., and Thompson, S. B. (2007). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Rev. Financ. Stud., 21*(4), 1509-1531.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance, 52*(1), 57-82.

Cumby, R. E., and Modest, D. M. (1987). Testing for market timing ability. *Journal of Financial Economics, 19*(1), 169-189.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy, 98*(4), 703-738.

Derrien, F. (2005). IPO Pricing in "Hot" Market Conditions: Who Leaves Money on the Table? *The Journal of Finance, 60*(1), 487-521.

Elton, E. J., Gruber, M. J., and Busse, J. A. (1998). Do Investors Care about Sentiment? *The Journal of Business, 71*(4), 477-500.

Fama, E. F., and French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics, 22*(1), 3-25.

Fama, E. F., and French, K. R. (1996). Multifactor Explanations of Asset Pricing Anomalies. *The Journal of Finance, 51*(1), 55-84.

Fama, E. F., and MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy, 81*(3), 607-636.

Fisher, K. L., and Statman, M. (2003). Consumer Confidence and Stock Returns. *The Journal of Portfolio Management, 30*(1), 115-127.

Goetzmann, W. N., Jr, J. I., and Ivkovic, Z. (2000). Monthly Measurement of Daily Timers. *The Journal of Financial and Quantitative Analysis, 35*(3), 257-290.

Henriksson, R. D., and Merton, R. C. (1981). On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills. *The Journal of Business, 54*(4), 513-533.

Hirshleifer, D. (2001). Investor Psychology and Asset Pricing. *The Journal of Finance, 56*(4), 1533-1597.

Lee, C. M. C., Shleifer, A., and Thaler, R. H. (1991). Investor Sentiment and the Closed-End Fund Puzzle. *The Journal of Finance, 46*(1), 75-109.

Lee, W. Y., Jiang, C. X., and Indro, D. C. (2002). Stock market volatility, excess returns, and the role of investor sentiment. *Journal of Banking and Finance*, *26*(12), 2277-2299.

Lemmon, M., and Portniaguina, E. (2006). Consumer Confidence and Asset Prices: Some Empirical Evidence. *Rev. Financ. Stud.*, *19*(4), 1499-1529.

Lettau, M., and Ludvigson, S. (2001). Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying. *Journal of Political Economy, 109*(6), 1238-1287.

Lowry, M. (2003). Why does IPO volume fluctuate so much? *Journal of Financial Economics*, *67*(1), 3-40.

Neal, R., and Wheatley, S. M. (1998). Do Measures of Investor Sentiment Predict Returns? *The Journal of Financial and Quantitative Analysis, 33*(4), 523-547.

Newey, W. K., and West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica, 55*(3), 703-708.

Pesaran, M. H., and Timmermann, A. (1994). Forecasting stock returns an examination of stock market trading in the presence of transaction costs. *Journal of Forecasting*, *13*(4), 335-367.

Schmeling, M. (2009). Investor sentiment and stock returns: Some international evidence. *Journal of Empirical Finance*, *16*(3), 394-408.

Statman, M., Thorley, S., and Vorkink, K. (2006). Investor Overconfidence and Trading Volume. *Rev. Financ. Stud.*, *19*(4), 1531-1565.

Treynor, J., and Mazuy, K. (1966). Can mutual funds outguess the market. *Harvard business review*, *44*(4), 131-136.

Table 2.1. Descriptive Statistics.

Statistic	RF	MKTRET	CC	BW	BWA
Mean	0.004560	0.010268	86.18487	0.260882	0.245618
Median	0.004242	0.014399	90.2	0.1135	0.054
Maximum	0.013583	0.12967	112	2.321	2.497
Minimum	0.000025	−0.219503	51.7	−1.333	−1.273
Std. Dev.	0.002836	0.044035	13.18374	0.677739	0.68605
CV	0.621882	4.288566	0.152970	2.597876	2.793166
Skewness	0.666981	−0.778491	−0.436221	0.765964	0.861307
Kurtosis	0.465557	5.695326	2.408766	3.624503	3.405068

Sample: July 1978–December 2010, monthly intervals, giving 390 observations of each variable. CV is the coefficient of variation, which allows for comparison of CC and BW; RF is the risk-free rate in decimal form, as measured by the 3-month T-Bill, obtained from the FRED; MKTRET is the CRSP value-weighted portfolio; CC is the University of Michigan Consumer Sentiment Index; BW is the Baker-Wurgler investor sentiment index, from Baker and Wurgler (2006), obtained from Jeffrey Wurgler's website; BWA is the Baker-Wurgler alternative index, which removes business cycle variation, from Baker and Wurgler (2006).

Table 2.2. FRS Descriptive Statistics.

Statistic	FRS
Mean	−0.319224
Median	−0.398
Maximum	1.815
Minimum	−1.254
Std. Dev.	0.53787
Skewness	1.452870
Kurtosis	6.435207

Sample is from April 1989 to November 2010. Thus, when forecasting is performed starting at the one-third mark of the full sample, excess market return is forecasted starting at May 1989 up until December 2010, at one-month intervals. FRS (fundamental-removed sentiment0 is the series of residuals from Equation (2.2), where only the final residual of each recursive regression is saved. This produces the time t residual, e, which is used in forecasting the market
at $t + 1$.

Table 2.3. Forecast Regression Results.

EMR:	**Model I**	**Model II**	**Model III**
BW			
Coefficient	−0.0036		
T-stat	(−1.91) *		
BWA			
Coefficient		−0.0037	
T-stat		(−1.73) *	
FRS			
Coefficient			−0.0040
T-stat			(−2.00) *
R-Squared	0.0041	0.0046	0.0044

* denotes significance at the 10% level. The table shows the average of coefficients from forecasting equations for various forecast models, with monthly market returns as the dependent variable. T-statistics are created using the Fama-MacBeth (1973) methodology. Forecasting starts at the halfway point of the sample (October 1994), rolling forward each month, creating 195 forecasting equations. Note that constants are not reported here. Model I uses the Baker-Wurgler index to forecast excess market returns, Model II uses the orthogonal Baker-Wurgler index that removes business-cycle variation to forecast, and Model III uses the new FRS (fundamental-removed sentiment) index from Equations (2.1)–(2.3). Sample: July 1978–December 2010, monthly intervals.

Table 2.4. Forecasting results.

Panel A: Realized Returns					
Time Series Statistic	**Model I**	**Model II**	**Model III**	**Model III - Model I**	**Model III - Model II**
Realized investment excess returns (average)	5.297%	7.978% **	10.643% ***	5.376% **	2.665% *
Realized investment excess returns (median)	1.807%	1.876%	9.394% ***	7.587% ***	7.518% ***
Realized returns after removing top 1% (average)	4.137%	6.846% **	9.539% ***	5.402% **	2.693% *
Standard deviation of realized returns	3.536%	3.216%	3.805%		
Realized returns after market risk correction	1.718%	4.967% ***	6.400% ***	4.682% **	1.433%
Realized returns after FF3 correction	2.283%	4.418% **	5.846% ***	3.563% *	1.428%
Realized returns after FF3 and momentum correction	1.571%	4.091% **	5.371% ***	3.800% *	1.280%

Table 2.4 continued

Panel B: Market Timing Ability Tests for Model III

Independent Variable	TM	HM	CM
α	0.0002 (0.002)	−0.0043 (0.003)	−0.0172 (0.009) *
R	0.7104 (0.087) ***	0.9244 (0.095) ***	
R^2	2.0954 (1.010) *		
D		0.4676 (0.221) **	
I			0.0283 (0.009) ***

Panel A shows the average annualized realized returns for the various models. Excess market returns use value-weighted CRSP market returns and 3-month T-bill rates for the risk-free rate. Realized investment returns are based off an investment strategy of investing in the market if predicted returns are higher than a stepwise average of excess market returns for all previous months or investing in the risk-free asset if predicted returns are lower than this average. Realized returns are from October 1994–December 2010. The third row shows the average returns when removing the top 1% of realized returns obtained. The realized returns are corrected for risk using a simple beta formulation, Fama and French (1996) 3-factor model, and Carhart (1997) 4-factor model. In columns 2-4, * represents significance at ten percent, ** at 5%, and *** at 1% using 20,000 simulations where the market and risk-free asset are randomly picked each month using the same ex-post weight of the corresponding model. See Section 2.4.3 for more details. Columns 5 and 6 in Panel A show the difference in returns between Models III and I, and Models III and II, respectively. The significance levels here are obtained from a bootstrapping procedure, using the sample of 195 return observations and 10,000 simulations. The null hypothesis is that the two series have a difference of zero. Panel B shows the market timing test results for Model III over the same sample period as Panel A. TM is the Treynor and Mazuy (1966) test, HM is the Henriksson and Merton (1981) test, and CM is the Cumby and Modest (1987) test. α is the intercept, R is the CRSP value-weighted excess market return, D is the dummy variable from Henriksson and Merton (1981): D = max[0, rf - r] (where rf is the return on 3-month U.S. Treasury bills), r is the excess market return, and I is the Cumby and Modest (1987) dummy variable: I = 1 if the model predicts a return higher than average and I = 0 otherwise. The Cumby and Modest (1987) test uses the excess market returns as the dependent variable. Coefficients for each variable are provided, along with the Newey-West (1987) standard errors in parentheses, with 2 lags based on Akaike Information criterion. * represents significance at the 10% level, ** significance at the 5% level, and *** significance at the 1% level.

Table 2.5. Realized returns after transaction costs.

Transaction Costs	High	Medium	Low
Model I	5.243%	5.268%	5.286%
Model II	7.901%	7.937%	7.963%
Model III	10.529%	10.585%	10.624%

This table shows the average realized excess returns for each forecast model. Model I uses the Baker-Wurgler (2006) index (BW), Model II uses the alternate Baker-Wurgler index that removes business cycle variation (BWA), and Model III uses fundamental-removed sentiment (FRS) to forecast. High transaction costs are 1% for stocks and 0.1% for bonds; medium costs are 0.5% for stocks and 0.1% for bonds; low costs are 0.15% for stocks and 0.1% for bonds. Transaction costs are subtracted from monthly returns depending on whether stocks (market) or bonds (risk-free asset/T-bills) are chosen for that particular month. For example, if stocks are chosen in month t, then the transaction-cost-adjusted return for t is r(t) x (1-c), where r is the return and c is the appropriate cost for stocks. Similar analysis is done for bonds, then an average of the full sample (October 1994–December 2010) adjusted returns is reported.

Table 2.6. Forecasting results for larger sample.

May 1989–December 2010 Forecasting Sample					
Time Series Statistic	**Model I**	**Model II**	**Model III**	**Model III - Model I**	**Model III - Model II**
Realized investment excess returns	5.192%	6.571% *	9.505% ***	4.313% **	2.934% **
Realized investment excess returns (median)	1.429%	2.039%	9.310% ***	7.881% ***	7.271% ***
Realized returns after removing top 1% (average)	3.860%	5.255% *	8.223% **	4.363% **	2.968% **
Standard deviation of realized returns	3.513%	3.182%	3.745%		
Realized returns after risk correction	1.279%	3.320% **	4.978% ***	3.699% **	1.658%
Realized returns after FF3 correction	0.117%	2.415% *	4.549% ***	4.432% ***	2.134% *
Realized returns after FF3 and momentum correction	−0.012%	2.083%	3.938% ***	3.950% **	1.855%

The above returns are annualized monthly returns. Excess market returns use the CRSP value-weighted average for market returns and the 3-month T-bill rate for the risk-free rate. Realized investment returns are based off investment strategy of investing in the market if predicted returns are higher than a stepwise average of excess market returns for all previous months or investing in the risk-free asset if predicted returns are lower than this average. Realized returns are from May 1989–December 2010. These realized returns are corrected for risk using a simple beta formulation, Fama and French (1996) 3-factor model, and Carhart (1997) 4-factor model. Model I uses the Baker-Wurgler (2006) index to forecast excess market returns, Model II uses the orthogonal Baker-Wurgler index that removes business-cycle variation to forecast, and Model III uses the new fundamental-removed sentiment (FRS) index from Equations (2.1)–(2.3). In Columns 2–4, * represents significance at ten percent, ** at 5%, and *** at 1% using 20,000 simulations where the market and risk-free asset are randomly picked each month using the same ex-post weight of the corresponding model. Again, the null hypothesis is that returns are not predictable. See Sections 2.4.3 and 2.4.4 for more details. Columns 5 and 6 show the difference in returns between Models III and I, and Models III and II, respectively. The significance levels here are obtained from a bootstrapping procedure, using the sample of 260 return observations and 10,000 simulations.

Table 2.7. Forecasting results for size portfolio.

Time Series Statistic	**Model I**	**Model II**	**Model III**
Realized investment excess returns	8.588%	10.577% *	13.315% **
Standard deviation of realized returns	5.308%	4.247%	5.599%
Realized returns after market risk correction	3.600%	7.361% *	8.778% **
Realized returns after FF3 correction	0.280%	4.506%	5.888% **
Realized returns after FF3 and momentum correction	0.649%	5.224% *	5.430% *

This table shows the average annualized excess realized returns for the various models when forecasting the small-size portfolio. Ken French's size-sorted portfolios are used, with the smallest decile used as the size portfolio to be forecasted. Realized investment returns are based off an investment strategy of investing in the small-size portfolio if predicted returns are higher than the stepwise average for all previous months or investing in the risk-free asset if predicted returns are lower than this average. 3-month US T-Bills are used for the risk-free asset. Realized returns are averages from October 1994–December 2010. The estimation window starts in July 1978. The realized returns are corrected for risk using a simple beta formulation, Fama and French (1996) 3-factor model, and Carhart (1997) 4-factor model. * represents significance at ten percent, ** at 5%, and *** at 1% using 20,000 simulations where the size portfolio and risk-free asset are randomly picked each month using the same ex-post weight of the corresponding model. See Section 2.4.3 for more details.

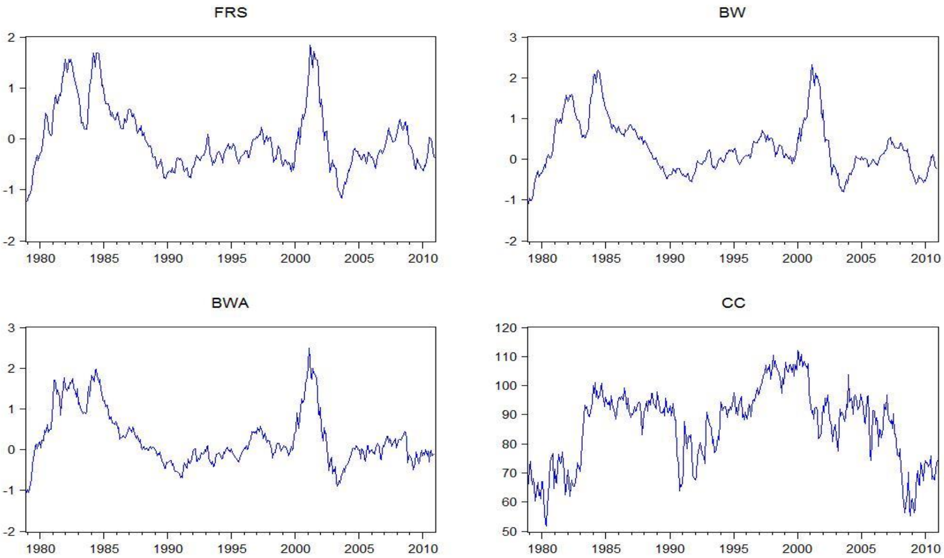


Figure 2.1: This figure shows time-series graphs of the paper's modified BW index (FRS), BW, BWA, and CC. FRS is fundamental-removed-sentiment, BW is Baker and Wurgler's (2006) sentiment index, BWA is Baker and Wurgler's sentiment index that removes business cycle variation, and CC is the University of Michigan Consumer Sentiment Index. The sample is from July 1978–December 2010.

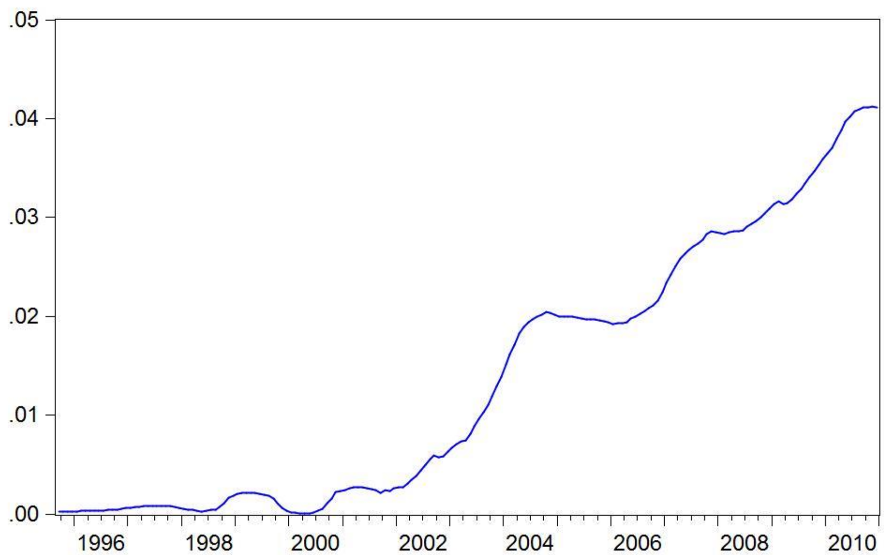


Figure 1.2 This figure shows the 12-month moving average R-squared values of the forecasting regression over the forecasting period. Here, forecasting is done with a rolling, 195-month regression in order to forecast the next month's return. The figure shows the R-squared values increasing over the forecasting period, which may explain the increasing performance of the market timing investment strategy.

Appendix 2.A

Table 2.A.1. Additional models' forecasting results.

Forecast Model	Realized Excess Investment Return
BW-fundamentals	3.725%
FRS and fit	1.770%
CC	2.172%
LPS	4.992%
BW-CC	5.255%
AR(1)	4.635%
AR(6)	2.276%
AR(12)	1.183%
PC	5.967%

This table presents the realized excess investment return that is obtained from the forecasting model using the appropriate variable. Excess market return is forecasted using the appropriate variable above. BW-fundamentals is the residual after removing the fundamentals used in this paper directly from the Baker and Wurgler (2006) index (BW), FRS and fit uses both fitted and residual components of Equation (2.2) in order to use both sentiment (FRS) and fundamentals (fitted value) in forecasting, CC is the University of Michigan Consumer Sentiment Index, LPS is the Lemmon and Portniaguina (2006) sentiment index (constructed by using the methodology in this paper: removing fundamentals from CC), BW-CC is the residual after removing the full CC index from BW, AR(p) is autoregressive model of order p, and PC is the residual after removing the first principal-component index of the fundamentals from BW.

Chapter Three: *Equity Premium Predictions and Optimal Hedging with Many Predictors Using Partial Least Squares*

3.1 Introduction

Two of the more persistent features in asset pricing are the size premium and value premium. The literature on these premiums provides either a rational or behavioral explanation. This paper offers a potential rational explanation for these premiums: they offer risk premiums due to the fact that small-size stocks/portfolios and value stocks/portfolios signal a change in future investment opportunities. A composite index of size and book-to-market sorted portfolio returns has significant forecasting power out of sample for the equity premium. The composite index places higher weights on small-size and high book-to-market ratio portfolios in the composite index. Therefore, small-size and value portfolios matter more in explaining future market returns. This signals a change in future investment opportunities: low returns on these portfolios mean future market returns will be low as well on average. Hence, these portfolios/stocks require a risk premium.

This paper argues that the composite index is an ICAPM state variable, since it reflects changes in future investment opportunities. Also, since the index places more weight on small-size and value portfolios compared to the market and a 1/N (equal-weighted) portfolio, it can be thought of as an economical proxy for the size and value factors of the Fama and French (1993) three-factor model. There is much literature that links forecasting future market returns and ICAPM state variables (and also showing that the Fama-French three-factor model can be considered an ICAPM). Campbell (1993) constructs an ICAPM which relates mean returns to their covariances with current and future market returns. In doing so, Campbell shows that any cross-sectional factors can only be included if they forecast future market returns (or labor income as human capital is included in this model). Thus, the composite index can be considered a Campbell (1993) factor. Campbell also mentions that the Fama-French three-factor model may be considered as an ICAPM in this setting. Cochrane (2005, p. 150) states that any variable (factor in this case) that has significant forecast power for the market return is inherently an ICAPM state variable. Liew and Vassalou (2000) show that the SMB (size) and HML (book-to-market) factors can predict future GDP growth in a cross-section of ten countries. They point out that this finding supports the hypothesis that the Fama-French three-factor model can be considered as an ICAPM, with the size and value premiums reflecting true risk factors. This is consistent with other literature on the premiums, such as Zhang (2005) and Petkova and Zhang (2005) in regards to the value premium, and Fama and French (2006) on the size and value premiums.

This paper contributes to the literature by providing a direct mechanism through which the size and book-to-market factors of the Fama-French three-factor model reflect future investment opportunities. These small-size and value stocks/portfolios are a leading market indicator (i.e., they predict the equity premium), hence the direct relationship to future investment opportunities. While

Liew and Vassalou (2000) show that the size and book-to-market factors can predict future GDP growth, this paper provides a more direct and market-based mechanism. Since this paper provides further evidence that the Fama-French three-factor model can be thought of as an ICAPM model, it also posits a potential explanation for the size and value premiums: they are ICAPM risk premiums as they reflect future investment opportunities.

The composite index used in forecasting is formed by employing a shrinkage estimator known as partial least squares (PLS), which was recently introduced to the Finance and Economics field by Kelly and Pruitt (2013). PLS is *a priori* a strong candidate as a forecast method because it estimates jointly those factors and factor sensitivities that in-sample provide the highest covariance with the forecast variable. The forecasting literature has found that shrinkage estimators work much better than OLS when using multiple predictors. OLS becomes noisy and does not typically work well when using multiple predictors, as the weights put on the predictors become extreme due to over-fitting. Shrinkage estimators avoid this problem by putting more conservative weights on the predictors. Using a shrinkage estimator also allows for stable forecasts and forecast weights. This allows for proper examination of the weights and may potentially work well in applications.

Two out-of-sample (OOS) applications employing the portfolio weights are performed: applying the weights in real-time to invest in the factor portfolio, and using the weights to hedge against market risk. The factor is formed by applying the weights (on an out-of-sample basis) that are obtained from each month's forecasting regression. The PLS factor outperforms a naïve equal-weighted approach and the market portfolio in terms of the OOS-Sharpe ratio. There is a strand of literature which finds that a naïve 1/N diversification strategy typically dominates all mean-variance and asset pricing techniques in terms of out-of-sample Sharpe ratios. DeMiguel et al. (2009) and Tu and Zhou (2011) are among those who find that this is indeed the case. The analysis in this paper explains why equal-weighted portfolios perform well out of sample: they load more heavily on size and value stocks compared to a value-weighted market measure, and these size and value stocks reflect future investment opportunities and hence ICAPM state variable risk.

In the second application of the weights, a risk-hedging approach is found to reduce overall risk and improve Sharpe ratios compared to the market portfolio. The mean-variance investor in this scenario would invest in the market, the risk-free asset (T-Bills), and the PLS factor as a zero-investment hedge. The weights of each are chosen out of sample by means of an optimization/risk minimization approach. This gives credit to the hypothesis that the PLS factor signals a change in future investment opportunities as it hedges well against future market risk.

The encouraging performance of the out-of-sample applications is to a large degree due to the fact that PLS provides both a more accurate and more stable forecast compared to OLS. This is particularly useful in the risk-hedging application. While the market is extremely volatile, the PLS weights are stable and forecast the market well enough, which allows an investor to hedge month-to-month without significant rebalancing. It is shown in this paper that OLS coefficients are more volatile over time, and the OOS performance of OLS is inferior to both the historical mean and PLS. If an investor were to use the OLS weights to construct a portfolio or hedge against market risk, this would require constant rebalancing and short-selling. In contrast, PLS produces positive, closely clustered, and stable weights that would be more useful in practice (in addition to producing superior forecasts). This paper examines the weights on the size- and value-sorted portfolios as well as on various industry portfolios that constitute the forecast factor.

This paper forecasts the equity premium rather than raw market returns, following many others in the forecasting literature. For example, Li et al. (2013) create an implied cost of capital measure to forecast the equity premium and Pettenuzzo et al. (2014) impose economic constraints to improve univariate forecasts of the equity premium. This paper borrows some of the out-of-sample tests from Li et al. A utility gain measure is implemented, which is equivalent to the certainty equivalent that an investor would pay to receive the forecast information.

The methodology implemented in this paper is closely related to that of Kelly and Pruitt (2013) and Kelly and Pruitt (2015). While this paper uses PLS to forecast, Kelly and Pruitt (2013) use a three-pass regression filter (3PRF). PLS and 3PRF are similar, but the differences (discussed below) do result in improved forecasting. As will be shown, PLS is more robust than 3PRF. However, it is important to note that the differences are small and the goal of this study is not to simply outperform 3PRF but instead to extend this methodology to another application and gain new insights. The methodology in this paper also differs from Kelly and Pruitt (2013) in terms of the predictors used. This paper uses portfolio returns, while Kelly and Pruitt use book-to-market ratios. Forecasting with returns rather than book-to-market ratios allows for the portfolio weights to be directly useful for practical applications. The PLS portfolio factors can be formed directly from the portfolio weights.

In addition to using size and book-to-market sorted portfolios as the base assets in forecasting the market, various industry portfolios are considered as well. This analysis allows for examination of which industries matter for future returns. Although this paper looks at several different Fama-French industry classifications, a clear pattern develops as to which industries have a large weight and which have a low weight. For example, the energy industry (mining, coal, oil and gas) and the telecommunications industry tend to have low weight when forecasting excess market returns. At the other end of the spectrum, the entertainment industry and finance industry have larger weights than other industries.

The sample period in this study is January 1930 – December 2010, with forecasting starting January 1980. Excess log returns are used, the CRSP value-weighted index is used for market returns, and the yield on 3-month Treasury bills is used for the risk-free rate. This paper looks at the 6, 25, and 100 size-value portfolios, as well as various Fama-French industry portfolios.

The paper proceeds as follows. Section 3.2 reviews the relevant literature. Section 3.3 introduces the partial least squares regression approach. Section 3.4 reviews the methodology and data. Section 3.5 provides the forecast results, examines the portfolio weights, looks at two applications of the weights, and conducts robustness checks. Section 3.6 concludes the paper.

3.2. Literature Review

Kelly and Pruitt (2015) create a three-pass regression filter (3PRF), which can be used to forecast. They note that it is basically a special case of partial least squares. Kelly and Pruitt (2013) is the first to apply PLS to finance. PLS is essentially a shrinkage estimator. Since this paper's methodology closely follows Kelly and Pruitt (2015), their methodology will be explored here in more depth. The first pass of the 3PRF of Kelly and Pruitt is a time series regression where each predictor variable is the dependent variable and the proxies (or the variables to be forecasted themselves) are the regressors. The first pass coefficients are stored and become the regressors in the second pass cross-sectional regression where the predictors are again the dependent variable. The second pass coefficients are then used as predictors in a third pass predictive regression. The three passes can be summarized as follows:

1. $x_{i,t} = \phi_{0,i} + z_t^{'}\phi_i + \varepsilon_{i,t}$

2. $x_{i,t} = \phi_{0,t} + \hat{\phi}_i^{'} F_t + \varepsilon_{i,t}$

3. $y_{t+1} = \beta_o + \hat{F}_t^{'}\beta + \eta_{t+1}$

where x are the predictors, z is a proxy to be selected by the forecaster, and y is the variable to be forecasted. The first pass estimates how the predictors relate to the latent factors, and the second pass estimates the factors themselves. Kelly and Pruitt (2015) find that this approach is especially useful when using a large set of predictors. The 3PRF is shown to be the solution to a constrained least squares problem. They forecast macroeconomic variables and conduct Monte Carlo simulations to show that their method is superior to OLS and several variants of principal-components analysis.

Kelly and Pruitt (2013) apply the 3PRF to forecast market returns and cash flow growth on both a monthly and annual basis using portfolio book-to-market ratios. They do so using the 6, 25, and 100 size and book-to-market sorted Fama-French portfolios. They also do not use a proxy for market returns in the first pass; instead, they use market returns as the regressor in the first pass (y in place of z in 1 above). They obtain significantly positive in-sample and out-of-sample

R-squared values (OOS calculated based on the historical mean). The results hold for various robustness tests, and for both US and international data.

This paper forecasts the equity premium while making some changes to Kelly and Pruitt's (2013, 2015) methodology. The literature on predictability of the equity premium is vast and diverse. For a summary of approaches to forecasting the equity premium, see Welch and Goyal (2008) or Spiegel (2008). Pettenuzzo et al. (2014) find that implementing economic constraints improves the results when forecasting the equity premium. Li et al. (2013) create their own implied cost of capital measure, which they find has significant out-of-sample forecast power for the equity premium. Polk et al. (2006) use the cross-sectional price of risk to forecast the equity premium. Ferreira and Santa-Clara (2011) use a sum-of-parts method that combines the three components of market returns (dividend-price ratio, earnings growth, and price-earnings ratio growth). They find that this performs better than other variables as well as the three components individually. Rapach et al. (2010) combine individual forecasts to better predict the equity premium. Guo (2006) uses the consumption-to-wealth ratio (cay) of Lettau and Ludvigson (2001) to forecast market returns. Guo finds that this measure has substantial out-of-sample predictive power. This is in contrast to Bossaerts and Hillion (1999) and Goyal and Welch (2003), who find little out-of-sample predictive power of traditional forecast variables such as dividend yield, term premium, default premium, etc.

While this paper forecasts the equity premium through some changes to Kelly and Pruitt's (2013, 2015) methodology, the most important finding of this paper is a potential explanation for the size and value premiums. Various explanations have been offered for the size and value premiums. Zhang (2005) provides a model that can explain why value may be riskier than growth. Petkova and Zhang (2005) find that a conditional CAPM with time-varying risk can partially explain the value premium. They suggest APT or ICAPM-related risk factors to explain the value premium, which this paper explores. This line of thinking is supported by the theory of Campbell (1993) and the empirical results of Liew and Vassalou (2000). Fama and French (2006) find that the CAPM can explain the size and value premiums of 1928 to 1963, but not those of 1963 to 2004. Thus, current market risk cannot explain the size and value premiums. Israel and Moskowitz (2013) find that shorting does not explain a significant portion of the size and value premiums, and also that changes in transaction costs and markets over time do not affect the premiums. Maio and Santa-Clara (2015) find that dividend variation is more related to expected dividend growth rather than expected returns for size and value stocks, contrary to the result found for all stocks.

3.3. Partial least squares methodology

This paper finds that using PLS (specifically as applied via the SIMPLS algorithm) and portfolio returns as predictors delivers significant forecast performance based on OOS-R-squared values.[19] The de Jong SIMPLS algorithm for PLS is essentially as follows.[20] The matrix of predictors and the vector (in this case) of the target variable are multiplied and squared. Then eigenvalue decomposition is performed, where the first eigenvector is used to form the weights for the first factor. The weights and factors are then normalized. Note that the approach essentially amounts to finding the eigenvector/factor that maximizes the covariance between X and Y (predictors and target). This also allows for an equivalent one-step estimator that finds the factors and weights, as well the regression coefficients for each factor. The original squared matrix can be orthogonalized to the first factor so that the process can be repeated to find multiple factors if desired.

In more technical terms, the algorithm is as follows: the matrix Y of the variables to be forecasted is demeaned, and the dominant eigenvector of $X^T YY^T X$ selected. The eigenvector is post-multiplied by $X^T YY^T X$ to give the X factor weights, which weight the X matrix to generate a linear combination of the X variables that produces the X factor, or X-score. The X-score is normalized

[19] This paper uses SIMPLS instead of the 3PRF of Kelly and Pruitt (2013). However, Kelly and Pruitt (2013) note that their 3PRF reduces to PLS when using the forecast variable in the first pass instead of an alternative instrument, de-meaning and variance-standardizing the predictors prior to the first pass, and removing the constants from the first two passes. One other advantage of PLS compared to the 3PRF is that PLS (SIMPLS specifically) allows for more than one factor to be created (although this is not done in this paper). In addition to the NIPALS algorithm that is frequently used, de Jong (1993) creates a SIMPLS algorithm. De Jong notes that this algorithm avoids the construction of deflated data matrices while also calculating the factors directly from the variables. This allows for faster computation compared to NIPALS. Nonetheless, the SIMPLS algorithm is shown to produce basically identical results compared to NIPALS when there is only one dependent variable (as is the case in this paper). For these reasons, this paper follows the SIMPLS algorithm. For a summary of PLS, see Hoskuldsson (1988).

[20] For a full explanation and the detailed algorithm, see de Jong’s (1993) appendix.

to produce X loadings (X * X-score) and Y loadings (centered Y *X-score). The Y matrix times its loading gives the Y-score. The loadings are orthogonalized to all previous loadings and stored. Current loadings are then removed from the $X^T YY^T X$ matrix and the process is repeated for any additional factors. The PLS regression coefficients can also be conveniently found by multiplying the X weights by the Y loadings. PLS maximizes $Y^T XX^T Y$ or, equivalently, $X^T YY^T X$, which maximizes the covariance between Y and X (Y being the forecast variable(s) and X being the predictor variables in this case).

An additional advantage of PLS is that it provides a stable forecast over time, with diffuse weights.[21] OLS can give extreme weights due to over-fitting. Thus, when dealing with a large set of predictors, PLS should produce superior forecasts compared to OLS. PLS further allows for better examination of portfolio weights compared to OLS. The portfolio weights from PLS should be less volatile over the forecast period compared to OLS, as will be confirmed in Section 3.5. Stable weights and weights that are not as extreme as OLS would allow for an investor to better implement the information from the PLS forecasts. Knowing which portfolios matter for the next month's excess market return can allow the investor to hedge properly. It also would allow for a simple check of a few portfolios to generate a quick guess as to which direction the market should be headed in the next month.

[21] These are the properties of shrinkage estimators. PLS can be represented as a type of shrinkage estimator, as shown by Frank and Friedman (1993)

3.4. Data and methodology

Log excess monthly returns are used for all returns. This paper looks at the forecast power of the returns of Fama-French portfolios double-sorted on size (market capitalization) and book-to-market ratios.[22] The 6, 25, and 100 size-value portfolios will be examined as predictors. The CRSP value-weighted market average is used for the market portfolio, where excess market returns are the target measure (in excess of three-month T-bill rates). The sample period starts in January 1930 and ends December 2010 (sample period ends in 2010 in order to allow for direct comparison to the results of Kelly and Pruitt, 2013). Recursive forecasting starts in January 1980, to allow for at least a 50-year estimation window. All data are obtained from Ken French's website.

The de Jong SIMPLS algorithm (described in the previous section) is used to perform PLS for forecast purposes. Only the first PLS factor is used to forecast. The PLS regression is as follows:

$$(3.1) \quad Y_t = \alpha_t + B_t \mathbf{X}_{t-1} + \varepsilon_t ,$$

where B is a vector of PLS regression coefficients, X is a matrix of portfolio returns (which will be either T by 6, T by 25, or T by 100 depending on which set of portfolios used), Y is the vector of excess market returns, α is a constant, and ε is the error term. B*X gives the single X-factor (or X-score).

[22] These will also be referred to as size-value portfolios for brevity in what follows.

Note that Equation (3.1) has time subscripts since this is run recursively up to month t, in order to forecast month t+1. Then this in-sample estimation is performed up to month t+1 (with an anchored starting point of January 1930 and an expanding window) to then obtain a prediction for month t+2. Equation (3.1) gives the regression coefficients, which can alternatively be found directly from the SIMPLS algorithm:

$$B_t = \mathbf{R}_t Q_t^{\,T}, \tag{3.2}$$

where **R** is the matrix of X weights, and Q is the vector of Y loadings. Then the forecast for Y_{t+1} is created as follows in order to implement the X information up to time t:

$$\hat{Y}_{t+1} = X_t' \hat{B}_t, \tag{3.3}$$

where X is a vector containing portfolio returns of only month t, and $\hat{B}$ is the estimated vector of coefficients obtained either from Equation (3.1) or (3.2).

Repeating this recursively produces a series of predicted market returns, $\hat{Y}$ (or $\hat{r}$), and a series of realized values of Y (or r). The OOS-R^2 statistic based on the historical mean can be produced from here, as in Campbell and Thompson (2008). The OOS-R^2 is as follows:

$$OOSR^2 = 1 - \frac{\sum_{t=1}^{T} (\hat{r}_t - r_t)^2}{\sum_{t=1}^{T} (\bar{r}_t - r_t)^2}, \tag{3.4}$$

where T is the last month of the forecast sample, r is the market return, $\hat{r}$ is the forecast of the market, and $\bar{r}$ is the current historical average (at time t). Note that the OOS-R^2 statistic has a range of $(-\infty,1]$ and provides a measure of how much better the forecast is compared to the historical mean.

3.5. Forecast results and portfolio weights analysis

3.5.1 Forecast results

First, the forecast results are examined to test whether the model and composite portfolio return factor has predictive power for the equity premium. Once it has been shown that there is indeed significant forecasting power then additional analysis can be performed. Table 3.1 shows the in-sample and out-of-sample R-squared values for each of the test assets when forecasting the equity premium. The p-values for the OOS-R^2 are constructed based on the mean squared performance error (MSPE) F-statistic of Clark and West (2007), as in Li et al. (2013). As can be seen in Table 3.1, for all tests assets examined, the OOS-R^2 is positive at a 1% significance level. The methodology is robust for all test assets used. Little to no explanatory power is lost when moving out of sample, as can be seen by comparing the in-sample and out-of-sample R^2 statistics. In a few cases, most likely due to chance, the OOS-R^2 is actually higher than the IS-R^2 statistic. Table 3.1 shows that the 100 portfolios provide the highest in-sample and out-of-sample R^2, although there is not a wide range between the three different groups of test assets. The same pattern holds in Kelly and Pruitt (2013), which is expected when using PLS: better forecast performance when using a larger set of predictors.

[Table 3.1 goes here]

The OOS-R^2 values are statistically significant, but Cochrane (2005) shows a way to measure economic significance for a mean-variance investor. The Cochrane measure for an active investor's Sharpe ratio, $s^* = \sqrt{\frac{s_0^2 + R^2}{1 - R^2}}$, can be used along with the historical monthly Sharpe ratio of .108 from Campbell and Thompson (2008) to show the potential Sharpe ratio improvement for an investor using the forecast information presented here. From the results in Table 3.1 this potential Sharpe ratio improvement ranges from 15.1% to 15.6% for the 6, 25, and 100 portfolio returns, comparable but slightly higher than those obtained by Kelly and Pruitt (2013).

Table 3.1 also shows the potential utility gain that a mean-variance investor could obtain by utilizing the information of the appropriate forecast model. This could be seen as the certainty equivalent (financial gain) or the fee an investor would be willing to pay a fund manager. This value is obtained by following the methodology of Li et al. (2013). Based on forecasts of return and variance, a mean-variance investor will optimally choose the weight placed on stocks (the market) for time *t+1* as follows:

$$w_{i,t} = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{i,t+1}}{\hat{\sigma}_{t+1}^{2}}\right). \tag{3.5}$$

Here, *i* represents the forecast model used (could also be the forecast based on the historical average for $\hat{r}_{i,t+1}$), and γ represents the relative risk aversion level for the investor (assumed to be three as in Li et al.). For all forecast models, including the historical average, the forecasted variance is calculated from a ten-year rolling window.

These weights are applied out of sample, one-month ahead, and the mean return and variance can be found for the forecast sample. Then, utility for the particular model (indicated by subscript i) is

$$U_i = \mu_i - \frac{1}{2}\gamma\sigma_i^2 . \tag{3.6}$$

The utility level for the forecast using the 6, 25, or 100 portfolio returns are then compared to the utility level of the forecast utilizing only the historical mean return. The difference of the PLS forecast utility level over the historical mean forecast utility level is reported as the utility gain in Table 3.1. Note that this is annualized and reported in percentage form. The forecast using size-value portfolio returns and PLS provides a utility gain over the historical mean forecast of around 3.1% to 3.4%. This can be thought of as the annual portfolio management fee that a mean-variance investor would pay for this forecast.

Two alternative methodologies are also used to forecast: OLS and a 1/N approach. The 1/N approach puts a weight of 1/N on each portfolio to give the next month's predicted excess market return. The same test assets are used, and the results are shown in Table 3.2. Both approaches give a large negative OOS-R^2 (as expected from Campbell and Thompson, 2007). The results show that there is predictive value provided by the PLS approach. Panel A shows that OLS does work well on an in-sample basis, but has no out-of-sample power compared to the historical average, while also providing a negative utility gain/certainty equivalent. Panel B shows that the 1/N approach forecasts poorly out of sample and gives a lower in-sample fit compared to PLS. However, the utility gain is actually higher compared to the PLS forecasts. This shows a potential downside to the Li et al. (2013) approach: it is possible to choose the weights an investor will put on stocks and bonds well without actually forecasting the market well, as is explained below. The 1/N portfolio is less accurate than the historical average as a forecast (according to the OOS-R^2), but the 1/N forecast puts a more optimal weight on stocks compared to the historical mean (under the constraints applied to the weights).

[Table 3.2 goes here]

However, Table 3.3 shows that the 1/N forecast actually gets the direction of the equity premium right less often than the PLS forecast, for all three groups of test assets. Table 3.3 also shows that removing the restrictions on the weights applied to the market (which in Tables 3.1 and 3.2 force the weight to be between 0 and 1.5, inclusive) produces extremely large utility losses for the 1/N forecast. This applies for each constraint applied individually as well (no short selling and a 50% leverage limit). Conversely, the utility gain for the PLS forecast usually only slightly decreases or in some instances actually increases when removing one or both constraints. Finally, Table 3.3 also shows the percentage of time that one of the constraints is binding (which applies to the results in Tables 3.1 and 3.2). The 1/N forecast must have the constraint imposed on the weights around 91% of the time, compared to between 22% and 25% for the PLS forecast. Therefore, it appears that the greater utility gains for the 1/N forecast found in Table 3.2 are only due to imposing constraints on the weights. The poor OOS-R^2 and utility gains without constraints shows that there is little predictive value in the 1/N portfolio. However, it does get the direction right more than 50% of the time (but less often than the PLS forecast), and this combined with the constraints previously imposed on the weights can create a large positive utility gain. These results suggest that the times where the forecast suggests a large weight in the wrong direction are reduced by the constraints, while the times where it gets the direction right still have a relatively higher weight compared to the PLS forecast (but still less extreme due to the constraints).

[Table 3.3 goes here]

3.5.2 Raw portfolio weight analysis

For reasons discussed earlier, PLS gives portfolio weights that should provide superior information to OLS weights. The other benefit is that the forecast performance is better. With more explanatory power, the portfolio weights will be more useful. PLS weights are mostly uniform (as is the case with shrinkage estimators) and generally positive (no short selling required). Table 3.4 also shows that PLS weights are much less volatile over the forecast sample compared to OLS weights. Since a new estimation and forecast is conducted for each month, the set of betas/weights can be stored for each month to produce a panel of weights. The table shows the standard deviation of the weights for each of the test assets in the time series dimension, for both PLS and OLS. The standard deviation reported is an average of the 6, 25, or 100 standard deviations for each of the individual weights. The time series standard deviation for the OLS weights range from 1.4% to 4.4%, while the PLS standard deviation is always close to zero, the highest being 0.1%.

[Table 3.4 goes here]

With the PLS weights being much more stable, this would help an investor intending to utilize the information. If the investor wanted to re-estimate the weights each month for either rebalancing or creating a new forecast, then there would be little cost to doing so. If an average bid-ask spread of 10 basis points is assumed (which is the low end that Pesaran and Timmermann, 1994 use), then the average transaction cost of using the OLS and PLS weights can be examined. For all three groups of test assets, implementing the PLS weights would give an average transaction cost of less than 1 basis point annually. For the OLS case, implementing the weights would give an average transaction cost of 18 basis points annually for the 6 portfolios, 30 basis points for the 25 portfolios, and 2.9% for the 100 portfolios. While the transaction costs for the first two groups of test assets are reasonable, investing in the 100 portfolios using OLS weights would mean a high transaction cost. However, the PLS weights consistently give a very low transaction cost that is near zero. Also, since the weights change little over time (the results in Table 3.4 are over the forecast sample of 30 years), the investor or practitioner could simply estimate the weights once and continue to use them (perhaps re-estimating every 1-2 years).

Since the portfolio weights from PLS should be more reliable and/or stable than those from OLS, it allows for examination of which portfolios explain future market returns well. Table 3.5 shows the ratio of the PLS weight to an equal-weighted portfolio for small size, large size, high book-to-market, and low book-to-market portfolios for the most extreme sorts and for sorts above and below the median sort. Providing the weight relative to an equal-weighted strategy allows for comparison across multiple sorts and portfolios. For example, the six portfolios have two size quantiles but three book-to-market quantiles. This analysis allows for a check of whether small- or large-size portfolios (and growth or value portfolios) matter more for explaining future excess market returns.[23] The second column in each panel shows where the portfolios rank on average. All portfolios in the group of test assets are sorted based on weight, the portfolio with the highest weight ranked first. Then the average rank for each characteristic is found, so that higher average weight means a smaller rank number. The table shows the results for the 6, 25, and 100 size-value sorted portfolios.

Examination of the results in Table 3.5 shows that small-size portfolios carry a greater weight than large-size portfolios, and value (high book-to-market) portfolios carry a greater weight than growth (low book-to-market) portfolios. Although the ratios may not look significantly different, it is important to remember that weights are closely clustered in PLS; the difference between the highest-weighted portfolio and the lowest-weighted portfolio is roughly 5%, 2.5%, and 2% for the 6, 25, and 100 portfolios respectively. Thus, it can be seen that the smallest size portfolio has a markedly higher weight than the largest size portfolio in explaining future excess market returns for each group of test assets (in addition to having a higher weight than an equal-weighted portfolio). The same pattern holds for the intermediate size portfolios, as portfolios with size below the median have a higher weight than portfolios with size above the median (and also a ratio above 1 compared to equal weights). The closer link between small firm returns and future market returns may offer an explanation for why

[23] The weights in this case were calculated based on full-sample estimation (i.e., the last forecast estimation run), but since weights are stable over time, the results obtained by averaging the weights from each monthly forecast are not materially different.

small-firm stocks have higher returns. They are riskier because they signal a change in investment opportunities: low returns on small stocks are followed by low returns in the wider market. This also means that when forming the composite PLS factor portfolio that will be used shortly, more weight is put on small stocks.

For the set of six portfolios, the highest book-to-market portfolios (value) have a higher weight compared to the lowest book-to-market portfolios (growth). The value portfolios here also have a weight much higher than an equal-weighted portfolio. The gap closes for the set of 25 portfolios, though, and reverses for the set of 100 portfolios for the lowest and highest book-to-market sorted portfolios, while both have weights below an equal-weighted portfolio. However, when looking at book-to-market portfolios above and below the median sort, it is clear that a higher book-to-market ratio results in a higher weight for all groups of test assets. The portfolios with a book-to-market ratio above the median also have a higher weight compared to an equal-weighted portfolio. Thus, again the higher returns of value stocks may be explained in part by their higher risk in that low returns on value stocks are followed by low market returns. Also, the PLS factor will place a higher weight on the value portfolios than other portfolios.

[Table 3.5 goes here]

3.5.3 Principal component weight analysis

An additional test is performed to investigate whether the results presented so far are spurious. This will help avoid the brunt of the criticism of Novy-Marx (2014). Novy-Marx finds that returns can be spuriously predicted by random variables such as weather, presidential elections, etc. The forecast may be doing well here by simply putting random weights on the predictors rather than providing any actual predictive information. To test against this alternative hypothesis, the following is performed. Rather than running PLS on the raw portfolio returns, PCA is performed first. From there, the methodology for forecasting is the same as above. For example, with the 25 size-value portfolios, PCA is performed, repackaging the original 25 portfolios, so that now there are 25 factors instead of the 25 raw portfolios. Note that these factors are simply linear combinations of the original returns, representing jointly the same information. Therefore, using these factors to forecast with PLS will not change the explanatory power (R-squared values do not change). However, now the weights should not be closely clustered if there is indeed some predictive information contained in the test assets. The first few factors from PCA should contain most of the information from the portfolio returns. Other factors should have relatively little weight.

From Figure 3.1, it can be seen that this is indeed the case. The figure shows the absolute value of the PLS weights, where the test assets are the PCA factors. Figure 3.1 shows the case for the 25 PCA factors obtained from the 25 size-value portfolios. The same pattern holds for all other test assets, therefore the figures are omitted. If the results were indeed spurious, then the weights on the factors would be random. However, this is not the case here. The first factor seems to capture all the relevant information, as it is the only factor with substantial weight. For the 25 factors in Figure 3.1, the first factor has a relative weight of around 57% whereas of all other factors, no other factor has a relative weight above 7%.[24]

[Figure 3.1 goes here]

3.5.4 PLS factor out-of-sample performance

Thus far, evidence has been presented which shows that PLS weights are stable with no negative weights in addition to having significant forecast power. One simple way to see how the weights would work in a practical way is to examine the PLS factor portfolio out of sample. The PLS weights from Equations (3.1) or (3.2) can be applied out of sample to essentially invest in the PLS factor portfolio on a real-time basis. At each month, PLS estimation is performed and PLS weights are stored. These weights can be used by an investor to invest in the

[24] Other test assets place a relative weight on the first factor of up to 86%. The number of test assets/factors is crucial to this weight; the more factors there are, the more the total weight is spread out. A similar test is performed where the weights on each raw portfolio are randomly selected, with the weights constrained to be positive and summing to one. The OOS-R^2 values are always significantly negative for all of the test assets when the forecast weights are randomly chosen with 5,000 simulations performed. Note that this can be seen as an alternative placebo test compared to the PCA analysis in this section. These results are available upon request. PLS also allows for examination of which portfolios capture future market risk and therefore do well in forecasting (see previous subsection).

6, 25, or 100 size-value sorted portfolios. Table 3.6 shows the average annualized excess return and standard deviation of this out-of-sample factor portfolio, along with the annualized Sharpe ratio. The six size-value portfolios provide the highest realized return and Sharpe ratio. The 100 portfolios provide the second-highest Sharpe ratio of about 0.42 compared to 0.43 for the six portfolios and 0.41 for the 25. For comparison, the market portfolio (or a buy-and-hold the market strategy) provides an average annualized excess return of 5.5% and a Sharpe ratio of 0.34. Note that transaction costs are ignored here, but as previously mentioned the PLS weights would give very low transaction costs so the adjustment is safely ignored here. In addition, all weights are positive so there would be no short selling required (in contrast to using OLS weights).[25]

In addition to out-performing the market, these PLS factor portfolios also beat a naïve 1/N diversification strategy (or equal-weighted portfolio) out of sample. There is much literature that finds that an equal-weighted strategy provides a higher OOS-Sharpe ratio than most mean-variance strategies or asset-pricing models. See DeMiguel et al. (2009), Tu and Zhou (2011), and others for more. It can be seen from comparing the two panels in Table 3.6 that this PLS factor would provide an investor with a higher out-of-sample Sharpe ratio than a naïve equal-weighted approach. Based on the results of DeMiguel et al., the PLS factor will likely also outperform a variety of mean-variance approaches out of

[25] The OLS weights produce factors with much lower Sharpe ratios (unreported). Using OLS weights would also require much more rebalancing due to their higher variance (see Table 3.4).

sample. However, simply beating the equal-weighted portfolio out of sample is an accomplishment. While the PLS factor of the 25 size-value portfolios only improves on the 1/N Sharpe ratio by 1.7%, the factor of the 100 portfolios improves on the 1/N Sharpe ratio by 9.1% and the factor of six portfolios offers a 5.9% improvement. Although the PLS factor may not substantially outperform the 1/N portfolio in some cases, it should be pointed out that the 1/N performs well because it places relatively more weight on the small-size and value portfolios compared to the market. This was derived endogenously by the PLS analysis, where the performance is boosted somewhat by putting even more weight on the small-size and value portfolios compared to the equal-weighted approach.

[Table 3.6 goes here]

Note that the Sharpe ratios in Table 3.6 shows how investing in the PLS factor portfolio would perform in real time. It should be expected that returns for this factor are higher than the market, as more weight is put on the small-size and value portfolios. The PLS factor and the market factor capture different risks: the market captures current systematic risk while the PLS factor captures changes in future investment opportunities. The weights on the factor are chosen according to which portfolios explain more of the future equity premium. It is not clear from the ICAPM whether the market or this PLS factor should have a higher Sharpe ratio. However, the evidence suggests that the PLS factor (and by extension size and value factors) has a higher price of risk: more return is required per unit of risk for the PLS factor compared to the market factor. Further, this adds evidence to the hypothesis that the size and value premiums are due to these stocks signaling a change in future investment opportunities. The ICAPM state variable risk explanation is also in line with the findings of Zhang (2005), Petkova and Zhang (2005), and Fama and French (2006): the CAPM cannot explain the size and value premiums, and an APT or ICAPM may be necessary to properly explain the differences in mean return across assets.

3.5.5 Hedging with PLS weights

As noted previously, the PLS weights are obtained from forecasting the equity premium. Therefore, rather than using the weights to form the PLS factor and apply it out of sample, an investor could potentially use the PLS weights to hedge against market risk. Since the PLS weights are stable and forecast the market well, they should work well as a hedge out of sample. Since the PLS factor reflects future market risk, it should be able to hedge against this risk on a real-time basis. To test this, the following hedging strategy is implemented. Each month the market portfolio and the risk-free asset are combined based on weights from a risk minimization technique that also includes a zero-investment (by borrowing at the risk-free rate) PLS factor portfolio as a hedge. This is done on an out-of-sample basis to see the real-time performance. The goal of the optimization approach is to reduce variance while keeping return at a constant level, which here is picked to be the market return. Note that this optimization is essentially maximizing the Sharpe ratio as well. The optimization approach is as follows:

(3.7) $$\underbrace{Min}_{x} \; E_t[\lambda_t(R_{t+1}-\overline{R})+(1-\lambda_t)(r_{f,t+1}-\bar{r}_f)-x_t((F_{t+1}-r_{f,t})-(\overline{F}-\bar{r}_f))]^2$$

s.t. $\lambda_t\overline{R}+(1-\lambda_t)\bar{r}_f+x_t(\overline{F}-\bar{r}_f)=\bar{\lambda}=\overline{R}$,

where R is the market return, λ is the weight given to the market portfolio, r_f is the return on three-month T-Bills, x is the weight given to the zero-investment hedge, F is the PLS factor return, and $\bar{\lambda}$ is a constant which is chosen here to be the average market return.

The optimization is performed each month so that the weights can be found and applied one-month ahead on an out-of-sample basis. The constraint can be written as:

(3.8) $$\lambda_t = 1+x_t\left(\frac{\overline{F}-\bar{r}_f}{\overline{R}-\bar{r}_f}\right)\equiv 1+cx_t\,.$$

Then the optimization can be simplified to (dropping the risk-free asset from the variance and plugging in the constraint):

(3.9) $$\underbrace{Min}_{x} \;\; E_t[(1+cx_t)(R_{t+1}-\overline{R})-x_t(F_{t+1}-\overline{F})]^2\,.$$

Solving the optimization gives:

(3.10) $$x_t = \frac{-E_t[c(R_{t+1}-\overline{R})]^2 + E_t[(F_{t+1}-\overline{F})(R_{t+1}-\overline{R})]}{E_t[c(R_{t+1}-\overline{R})-(F_{t+1}-\overline{F})]^2} \equiv \frac{\sigma_{FR,t+1} - c\sigma^2_{R,t+1}}{\sigma^2_{Z,t+1}},$$

where $\sigma^2_{Z,t} \equiv c^2\sigma^2_{R,t} + \sigma^2_{F,t} - 2c\sigma_{FF,t}$. Note that the x varies positively with the covariance between F and R (PLS factor and market). This means that, as the two become more correlated, the PLS hedging portfolio will be shorted with a higher weight. Conversely, as the market variance increases, less weight will be put on the hedging portfolio. This is due to the fact that the market cannot be forecasted as well when it is more volatile, and the PLS hedging portfolio will not approximate the market as well. Finally, when the PLS factor portfolio variance increases, less weight is put on the hedging portfolio itself. If the variance increases and the covariance with the market does not increase by the same proportion, then the PLS factor will not be as reliable as a hedge. Therefore, less weight will be put on the hedging portfolio as it becomes riskier.

Now that *x* has been obtained, it can be found each month in a recursive manner to be applied to the next month, and λ=1+*cx* can also be found each month to be applied to the following month in real time. Therefore, for each upcoming month the investor can invest λ in the market portfolio, 1-λ in T-Bills, and *x* in the zero-investment PLS factor portfolio (F-r_f). From Equation (3.10), it can be seen that x will be negative if the covariance between the factor portfolio and the market is less than the constant c times the variance of the market. If x is negative, then Equation (3.7) shows that the factor portfolio will be held with positive weight (after borrowing x at the risk-free rate to ensure zero investment). If the covariance between the factor and market is greater than c times the variance of the market, then x will be positive and this portfolio will be shorted, or held in negative weight.

The values of x and λ can be found empirically as well. Using the full estimation window (or last forecasting regression) gives values of x = -0.62 and λ = 0.19 for the 100 portfolios, x = -0.71 and λ = 0.05 for the 25 portfolios, and x = -1.05 and λ = -0.71 for the 6 portfolios. This means that the PLS factor is held with a fairly large positive weight, while the market factor has a relatively small weight with the rest put on the risk-free asset (for the six portfolios the market is actually shorted). The average of each of these x and λ values from all months (as the minimization problem is repeated each month of the forecasting sample) are qualitatively similar to those from the full estimation window. Examining each month's x and λ values shows that x is typically negative but sometimes positive (although the average is similar to the values reported previously), while lambda can vary substantially over the sample.

Table 3.7 shows the results of this out-of-sample optimization technique/hedging strategy for each of the 6, 25, and 100 size-value portfolios. The average annualized excess return, standard deviation, and annualized Sharpe ratios are reported. By looking at the Sharpe ratios, the six size-value portfolios work best, with a Sharpe ratio of about 0.4. The 25 portfolios produce a Sharpe ratio of about 0.35, and the 100 portfolios produce a Sharpe ratio of about 0.36. For comparison, the market portfolio is also included in the table. The market portfolio has a Sharpe ratio over this sample of about 0.34. Note that while the optimization sets mean return equal to the average market return, this is out-of-sample analysis so the realizations will not necessarily produce the same average return as the market. However, in all cases the risk (standard deviation) is reduced compared to the market. The hedging strategy for the six portfolios comes close to eliminating half of the market risk, while improving the Sharpe ratio by almost 19% compared to the market. While the 25 and 100 portfolios do not offer quite the same risk reduction, the hedging strategy would still improve Sharpe ratios compared to the market by about 4.5% and 6.3%, respectively. Thus, the PLS factor does reduce risk and hedge against future market risk, meaning it does reflect future investment opportunities, as hypothesized.

[Table 3.7 goes here]

3.5.6 Robustness tests

To see how robust the forecasting results are and to gain other insights, various Fama-French industry-sorted portfolio returns are used as well. The 5, 10, 12, 17, 30, and 49 industry portfolios are used to forecast the excess market return using PLS. The results of this are shown in Table 3.8. It can be seen that the OOS-R^2 values are again significant at 1% compared to the historical market return average. In some cases, the OOS-R^2 values are even higher than those obtained by using the size-value sorted portfolios. The in-sample R^2 values are qualitatively similar to both the OOS-R^2 values here and to the IS-R^2 values from the size-value portfolios. The utility gain/certainty equivalents that these forecasts would offer to an investor are slightly lower than those offered by the size-value portfolios, yet still positive. It is clear from Table 3.8 that the forecast power of portfolio returns using PLS is robust to various portfolio sorts.

[Table 3.8 goes here]

The weights on the industry portfolios can be examined to see which industries matter in explaining the next month's excess market return. One issue is that the number of industries specified is crucial to the results. However, a few general results can be gleaned from examining the weights for all the different industry portfolios.[26] The energy industry (oil, mines, coal in some specifications) consistently has a low weight. The energy industry typically has a high idiosyncratic variance, which may explain the low weight on future market returns. Other industries that have a low weight across all specifications include telecommunications, food, soda, and clothing. These industries typically have stable returns even if investment opportunities change, so there is no risk premium or predictive value in these industries. The finance/money industry consistently has a high weight, perhaps suggesting a drawback to dropping these firms when collecting data. Note also that the financial sector has a lot to gain

[26] The industry weights are unreported, but available upon request.

from improving market conditions and thus it makes sense that the stock returns in this sector are useful leading indicators. In two of the seven specifications, the entertainment industry has the highest weights. Durables, paper, high-tech industries, books, automobile, and insurance also have high weights.

To generate a direct comparison to the results of Kelly and Pruitt (2013), raw market returns (non-excess) are also forecasted in the same manner as previously discussed for the equity premium. Table 3.9 shows the results for both the three groups of size-value sorted test assets and the industry-sorted portfolios. The assets still show significant forecast power for all cases. The values are also higher across the board compared to the results of Kelly and Pruitt (2013). Kelly and Pruitt obtain an OOS-R^2 of 0.65%, 0.77%, and 0.93% when using the book-to-market ratios of the 6, 25, and 100 size-value portfolios, respectively.[27] Kelly and Pruitt do not use an alternative instrument in the first pass, so the only differences between their methodology and PLS is normalization and constants in the first two passes. By making these changes one at a time, it appears that dropping the constants is what results in improvement. However, returns still work better than book-to-market ratios, so the improvement is not solely due to the difference in methodology.

In comparing the results in this table to Tables 3.1 and 3.8, it can be seen that the OOS-R^2 values are a bit lower across the board when forecasting market

[27] They do obtain much higher OOS-R^2 values [than you? Explain] compared to their monthly results when forecasting with annual returns. These results are available upon request.

returns rather than the equity premium. However, the PLS methodology is still more robust than 3PRF. 3PRF does not perform well when using returns, with negative OOS-R^2 values for all three groups of test assets used in Kelly and Pruitt (2013). Also, using 3PRF and book-to-market ratios as in Kelly and Pruitt does not work well when forecasting the equity premium, with positive but insignificant OOS-R^2 values.[28]

[Table 3.9 goes here]

3.6. Conclusion

The main contribution of this paper is to offer a potential explanation for the size and value premiums: they offer risk premiums as these stocks are leading indicators of future market returns and thus investment opportunities. A forecast factor is introduced, which is designed to explain future equity premiums. This factor places higher weights on the small-size and value portfolios, as these explain more of the future equity premium than large-size and growth portfolios. Forecasting is performed by partial least squares (PLS), which is a shrinkage estimator that also allows a single composite factor to be formed. This factor captures the variance of the predictor variables as well as the covariance with the market/equity premium.

The Sharpe ratio of the PLS factor portfolio is significantly different (higher) than the market factor. This is evidence that there are two different risk

[28] These results are unreported, but available upon request.

factors captured by each of these factors: current systematic risk and risk due to changes in future investment opportunities. Therefore, the PLS factor can be seen as capturing state variable risk as in Merton's (1973) ICAPM. Since the factor is constituted by more small-size and value portfolios than large-size and growth portfolios, then by extension small-size and value portfolios capture the risk due to changes in future investment opportunities. This is also evidence that the size and value factors from the Fama-French three-factor model are indeed true risk factors if the three-factor model is viewed as an ICAPM model. As has been suggested in previous literature, an additional risk factor over the market factor is necessary to explain these premiums. However, this paper provides a direct and observable mechanism by which these risk factors reflect changes in future investment opportunities.

This analysis is possible due to the PLS factor having significant forecast power for the equity premium. The out-of-sample-R^2 values compared to the historical mean are significantly positive. Also, it is shown that a potential investor would pay in excess of 3% of invested wealth annually in order to obtain the forecast information. Therefore, the results are both statistically and economically significant. These results are robust to using industry-sorted portfolios rather than size-value-sorted portfolios. Several industries are found to be consistently important in explaining future market returns: finance, durables, paper, high-tech industries, books, automobile, and insurance. Also, the methodology used in the paper offer weights that are more stable and less extreme than OLS weights. It is shown that this methodology works much better in forecasting compared to OLS when many predictors are used. These benefits allow the weights to be properly examined and a factor portfolio formed which can be examined as well.

To examine the PLS factor portfolio properties more closely, two applications are performed. The first application is to simply invest in the PLS factor out of sample. This is done by applying the PLS weights obtained by forecasting the market portfolio one-month ahead. While there is much research showing that a naïve equal-weighted diversification approach works better than most of the existing mean-variance approaches, this PLS factor outperforms the equal-weighted portfolio in terms of out-of-sample Sharpe ratio for all size-value sorted portfolios used in this paper. The PLS factor also outperforms the market portfolio in this regard. It is argued that both the PLS factor and the equal-weighted portfolio perform well due to the fact that they load more heavily on small-size and value compared to the market portfolio. The PLS factor performs somewhat better than the equal-weighted portfolio as it places a higher weight on small and value portfolios compared to the equal-weighted portfolio. These results give a possible explanation for why the equal-weighted approach typically works well: it loads more heavily on the ICAPM state variable risk of size and value.

The second application of the PLS weights is to conduct a risk-minimization/optimization approach in order to hedge against market risk. This is done by using the PLS factor (formed by borrowing at the risk-free rate) as a zero-investment hedge, and choosing optimal weights on both the market and the risk-free asset. Since the weights are obtained by forecasting and there is significant forecast power, these weights act as a natural hedge to the market. For all three of the size-value sorts, overall risk is reduced and the Sharpe ratio is improved on an out-of-sample basis. These two applications add to the evidence that the factor indeed captures a risk different from that of the market: a change in future investment opportunities.

Another contribution of the paper is to address a concern raised by Novy-Marx (2014): is the predictive power due to beneficial information or is it simply due to chance? By using PCA factors as predictors, it is found that most of the PLS weight is placed on the first factor. If the results were random, then random (most likely diffuse) weights would be placed on all factors. Further, with only the first factor having significant weight, it is argued that the only risk that matters is a combination of market risk and the risk of changes in investment opportunities.

References

Bossaerts, P., and Hillion, P. (1999). Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn? *Rev. Financ. Stud., 12*(2), 405-428.

Campbell, J. Y. (1993). *Understanding risk and return*. National Bureau of Economic Research Working Paper No. w4554.

Campbell, J. Y., and Thompson, S. B. (2007). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Rev. Financ. Stud., 21*(4), 1509-1531.

Clark, T. E., and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics, 138*(1), 291-311.

Cochrane, J. H. (2005). *Asset Pricing*, Princeton University Press.

de Jong, S. (1993). SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems, 18*(3), 251-263.

DeMiguel, V., Garlappi, L., and Uppal, R. (2007). Optimal Versus Naive Diversification: How Inefficient is the 1/ N Portfolio Strategy? *Rev. Financ. Stud., 22*(5), 1915-1953.

Fama, E. F., and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3-56.

Fama, E. F., and French, K. R. (2006). The Value Premium and the CAPM. *The Journal of Finance, 61*(5), 2163-2185.

Ferreira, M. A., and Santa-Clara, P. (2011). Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics, 100*(3), 514-537.

Frank, l. E., and Friedman, J. H. (1993). A Statistical View of Some Chemometrics Regression Tools. *Technometrics, 35*(2), 109-135.

Goyal, A., and Welch, I. (2003). Predicting the Equity Premium with Dividend Ratios. *Management Science, 49*(5), 639-654.

Guo, H. (2006). On the Out-of-Sample Predictability of Stock Market Returns. *The Journal of Business, 79*(2), 645-670.

Höskuldsson, A. (1988). PLS regression methods. *J. Chemometrics, 2*(3), 211-228.

Israel, R., and Moskowitz, T. J. (2013). The role of shorting, firm size, and time on market anomalies. *Journal of Financial Economics, 108*(2), 275-301.

Kelly, B., and Pruitt, S. (2013). Market Expectations in the Cross-Section of Present Values. *The Journal of Finance*, *68*(5), 1721-1756.

Kelly, B., and Pruitt, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics, 186*(2), 294-316.

Lettau, M., and Ludvigson, S. (2001). Consumption, Aggregate Wealth, and Expected Stock Returns. *The Journal of Finance, 56*(3), 815-849.

Li, Y., Ng, D. T., and Swaminathan, B. (2013). Predicting market returns using aggregate implied cost of capital. *Journal of Financial Economics, 110*(2), 419-436.

Liew, J., and Vassalou, M. (2000). Can book-to-market, size and momentum be risk factors that predict economic growth? *Journal of Financial Economics, 57*(2), 221-245.

Maio, P., and Santa-Clara, P. (2015). Dividend Yields, Dividend Growth, and Return Predictability in the Cross Section of Stocks. *Journal of Financial and Quantitative Analysis*, *50*(1-2), 33-60.

Merton, R. C. (1973). An Intertemporal Capital Asset Pricing Model. *Econometrica*, *41*(5), 867-887.

Novy-Marx, R. (2014). Predicting anomaly performance with politics, the weather, global warming, sunspots, and the stars. *Journal of Financial Economics, 112*(2), 137-146.

Pesaran, M. H., and Timmermann, A. (1994). Forecasting stock returns an examination of stock market trading in the presence of transaction costs. *Journal of Forecasting, 13*(4), 335-367.

Petkova, R., and Zhang, L. (2005). Is value riskier than growth? *Journal of Financial Economics, 78*(1), 187-202.

Pettenuzzo, D., Timmermann, A., and Valkanov, R. (2014). Forecasting stock returns under economic constraints. *Journal of Financial Economics, 114*(3), 517-553.

Polk, C., Thompson, S., and Vuolteenaho, T. (2006). Cross-sectional forecasts of the equity premium. *Journal of Financial Economics, 81*(1), 101-141.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2009). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *Rev. Financ. Stud., 23*(2), 821-862.

Spiegel, M. (2008). Forecasting the Equity Premium: Where We Stand Today. *Rev. Financ. Stud., 21*(4), 1453-1454.

Tu, J., and Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics, 99*(1), 204-215.

Welch, I., and Goyal, A. (2007). A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. *Rev. Financ. Stud., 21*(4), 1455-1508.

Zhang, L. U. (2005). The Value Premium. *The Journal of Finance, 60*(1), 67-103.

Table 3.1: Excess market return forecast results

	6 SV	25 SV	100 SV
OOS R-squared	1.10%	1.17%	1.24%
OOS p-value	< 0.01	< 0.01	< 0.01
IS R-squared	1.18%	1.12%	1.34%
Utility gain	3.42%	3.09%	3.19%

This table shows the R-squared values from forecasting. Out-of-sample (OOS) R-squared is based on the historical mean of the market return. OOS p-value is the p-value obtained by regressing the Clark and West (2007) adjusted-MSPE statistic on a constant, as in Li et al. (2013). IS R-squared is the percent variance of Y (excess market returns) explained by the first X-component. Utility gain is the annualized average return that a mean-variance investor with a risk aversion level of three would gain (or be willing to pay; certainty equivalent) based on utilizing the forecasts compared to the historical mean.SV are Fama-French size-value sorted portfolio returns. Monthly data is used. Data is obtained from Ken French's website. Recursive forecasting is performed with the estimation sample starting at January 1930 (anchored starting point) to December 1979. Forecast sample starts at January 1980 and runs to December 2010. PLS regression estimates obtained from SIMPLS algorithm of de Jong (1993) are used to forecast. Excess log returns are used, the CRSP value-weighted market index is used for market returns, and 3-month US Treasury bills are used for the risk-free rate.

Table 3.2: Additional Equity Premium Forecast Results

	Panel A: OLS		
	6 SV	25 SV	100 SV
OOS R-squared	-3.17%	-17.68%	-28.82%
IS R-squared	1.67%	6.80%	11.75%
Utility gain	-0.22%	-0.49%	-2.53%
	Panel B: 1/N		
	6 SV	25 SV	100 SV
OOS R-squared	-105.52%	-92.79%	-94.73%
IS R-squared	0.68%	0.55%	0.55%
Utility gain	4.88%	4.39%	4.72%

This table shows the R-squared values from forecasting. Panel A uses OLS to forecast and Panel B uses the equal-weighted (1/N) portfolio to forecast. Out-of-sample (OOS) R-squared is based on the historical mean of the market return. Note from Equation (3.4) that this statistic can range to –infinity. IS R-squared is the adjusted R-squared value of the in-sample regression (equity premium on lagged portfolios). Utility gain is the annualized average return that a mean-variance investor with a risk aversion level of three would gain (or be willing to pay; certainty equivalent) based on utilizing the forecasts compared to the historical mean.SV are Fama-French size-value sorted portfolio returns. Monthly data is used. Data is obtained from Ken French's website. Recursive forecasting is performed with the estimation sample starting at January 1930 (anchored starting point) to December 1979. Forecast sample starts at January 1980 and runs to December 2010. Excess log returns are used, the CRSP value-weighted market index is used for market returns, and 3-month US Treasury bills are used for the risk-free rate.

Table 3.3: Utility Gain Without Constraints

	Panel A: PLS		
	6 SV	25 SV	100 SV
No short selling	2.44%	3.10%	3.31%
No 50% leverage limit	3.22%	3.06%	2.95%
Neither constraint	2.24%	3.07%	3.20%
Frequency of constraints binding	0.22	0.22	0.25
Directional accuracy of prediction	0.570	0.565	0.557
	Panel B: 1/N		
	6 SV	25 SV	100 SV
No short selling	-126.88%	-56.50%	-57.32%
No 50% leverage limit	-359.81%	-323.56%	-336.10%
Neither constraint	-490.16%	-383.08%	-396.71%
Frequency of constraints binding	0.92	0.91	0.91
Directional accuracy of prediction	0.535	0.548	0.546

This table shows the utility gain of implementing the forecast over using the historical mean as a forecast. See Equations (3.5) and (3.6) for details. Panel A shows the results when the PLS factor portfolio is used as the predictor, while Panel B shows the results when using the equal-weighted average of the test assets as the predictor. SV refers to the various size-value sorted Fama-French portfolios. The first two rows in each panel show the utility gain when relaxing the constraints of no short selling the market and a 50% leverage limit, respectively, individually. The third row shows the utility gain when relaxing both constraints. Utility gains are annualized. The fourth row shows the fraction of months when one of the constraints is binding. The last row shows the fraction of months when the prediction of excess market return is in the right direction of the actual excess market return.

Table 3.4: Comparison of OLS and PLS coefficient volatility

	Time series variance of coefficients		
	6 SV	25 SV	100 SV
OLS std dev	4.42%	2.53%	1.71%
PLS std dev	0.06%	0.02%	0.01%

This table compares, for each of the test assets, the average time series standard deviation of each of the OLS coefficients and PLS coefficients over the forecast sample period. Each time the recursive forecast regression is run, the coefficients for each portfolio are stored. Then for every portfolio's coefficients, the standard deviation is calculated. Reported in the table is the average standard deviation of the respective group of test assets. SV stands for size-value sorted portfolios (Fama-French portfolios). Recursive forecasting is performed with the estimation sample starting at January 1930 (anchored starting point) to December 1979. Forecast sample starts at January 1980 and runs to December 2010. PLS regression estimates obtained from SIMPLS algorithm of de Jong (1993) are used to forecast. Excess log returns are used, the CRSP value-weighted market index is used for market returns, and 3-month US Treasury bills are used for the risk-free rate.

Table 3.5: Summary of size and value weights

	6 size-value sorted portfolios		25 size-value sorted portfolios		100 size-value sorted portfolios	
	Weight/EW ratio	average rank	Weight/EW ratio	average rank	Weight/EW ratio	average rank
small size	1.02	3.3	1.07	10.2	1.06	47.1
large size	0.98	3.7	0.89	17.4	0.71	65.6
size below median			1.07		1.05	
size above median			0.90		0.95	
high BTM	1.12	1.5	0.99	13.6	0.85	64.7
low BTM	0.91	5.0	0.98	13.4	0.74	64.7
BTM above median			1.05		1.04	
BTM below median			0.94		0.96	

This table gives a summary of weights on various size and value portfolios, obtained from PLS regressions that forecast the market return. Weight/EW ratio is the ratio of the normalized (to sum to 1) weight of the portfolio(s) to its weight in an equal-weighted strategy. For example, small size for the 6 portfolios gives the ratio of the total weight put on the three small size to 0.5 (done for 6, 25, and 100 Fama-French size and book-to-market sorted portfolios, obtained from Ken French's website). Average rank shows where the average size or value sorted portfolio ranks in terms of weight compared to all the other portfolios. Recursive forecasting is performed with the estimation sample starting at January 1930 (anchored starting point) to December 1979. Forecast sample starts at January 1980 and runs to December 2010. PLS regression estimates obtained from SIMPLS algorithm of de Jong (1993) are used to forecast. Excess log returns are used, the CRSP value-weighted market index is used for market returns, and 3-month US Treasury bills are used for the risk-free rate.

Table 3.6: Out-of-sample performance of factor portfolio and equal-weighted portfolio

	PLS factor portfolio			Equal-weighted portfolio		
	Mean excess return	Standard deviation	Sharpe ratio	Mean excess return	Standard deviation	Sharpe ratio
6 SV	8.04%	5.36%	0.432	7.68%	5.41%	0.408
25 SV	7.20%	5.14%	0.401	6.96%	5.13%	0.394
100 SV	7.44%	5.13%	0.419	6.96%	5.19%	0.384

This table shows the mean excess return and the standard deviation for two different portfolios for each of the test assets. The first is the PLS factor portfolio, which is a portfolio constructed based on the portfolio weights obtained from the PLS regression. This is done out of sample. The second is the out-of-sample performance of a 1/n or equal-weighted approach. SV are Fama-French size-value sorted portfolio returns. Monthly data is used. Recursive forecasting is performed with the estimation sample starting at January 1930 (anchored starting point) to December 1979. Forecast sample starts at January 1980 and runs to December 2010. PLS regression estimates obtained from SIMPLS algorithm of de Jong (1993) are used to forecast. Excess log returns are used, the CRSP value-weighted market index is used for market returns, and 3-month US Treasury bills are used for the risk-free rate. Returns below are monthly, annualized excess returns.

Table 3.7: Out-of-sample risk hedging

	Mean excess return	Standard deviation	Sharpe ratio
6 SV	3.84%	2.82%	0.399
25 SV	4.80%	3.93%	0.351
100 SV	5.16%	4.18%	0.357
Market	5.52%	4.70%	0.336

This table shows the out-of-sample results of the risk minimization approach for each of the 6, 25, and 100 size-value (SV) sorted Fama-French portfolios, along with the market portfolio for comparison. See Equations (3.6)-(3.9) for the optimization approach. For each of the test assets, the corresponding PLS factor is used as a hedge. The market portfolio is combined with the risk-free asset (T-Bills) and the zero-investment PLS factor portfolio as a hedge. The mean excess returns reported are annualized, and the Sharpe ratios are annualized as well. The estimation window is from January 1930 - December 1979, and the out-of-sample tests start in January 1980 and extend to December 2010. Thus, the above statistics are calculated from January 1980 to December 2010.

Table 3.8: Forecast results for industry portfolios

	5 ind	10 ind	12 ind	17 ind	30 ind	49 ind
OOS R-squared	1.35%	1.25%	1.32%	1.11%	1.09%	1.12%
OOS p-value	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
IS R-squared	1.27%	1.29%	1.32%	1.18%	1.27%	1.31%
Utility gain	2.98%	2.58%	2.79%	2.25%	2.20%	2.46%

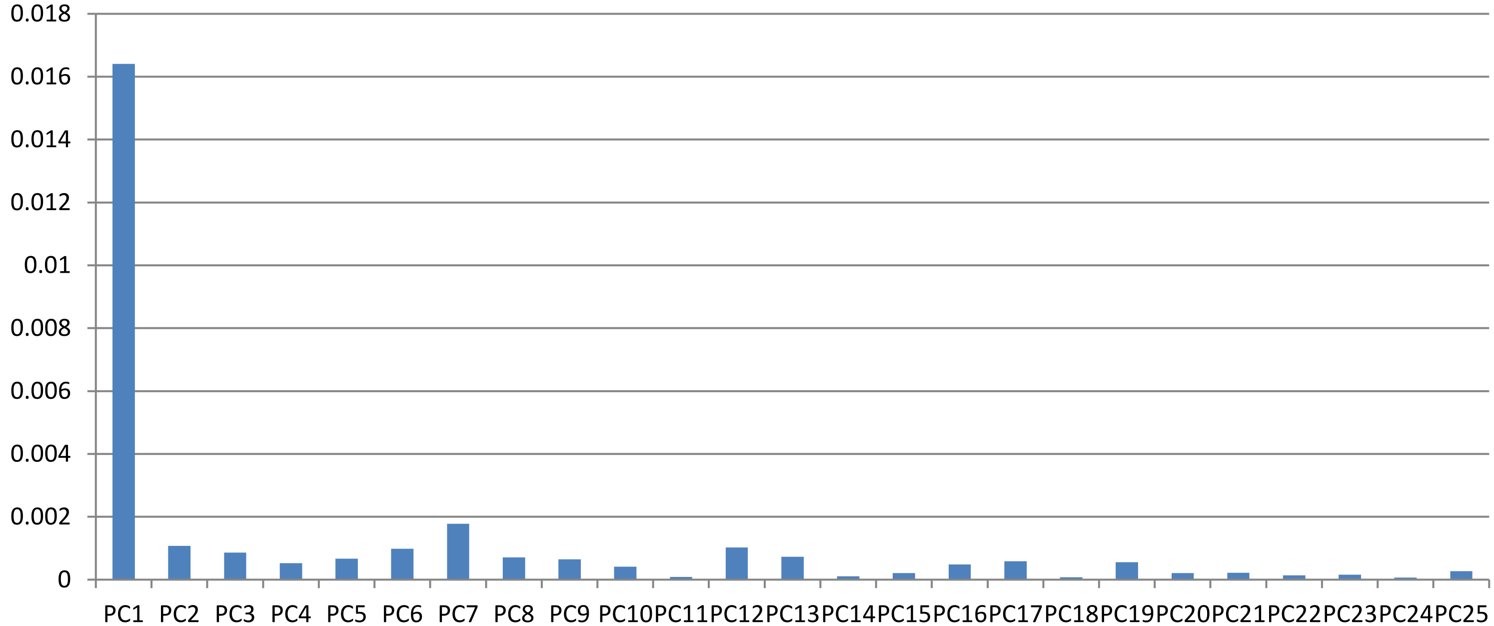
This table shows the R-squared values from forecasting. Out-of-sample (OOS) R-squared is based on the historical mean of the market return. OOS p-value is the p-value obtained by regressing the Clark and West (2007) adjusted-MSPE statistic on a constant, as in Li et al. (2013). IS R-squared is the percent variance of Y (excess market returns) explained by the first X-component. Utility gain is the annualized average return that a mean-variance investor with a risk aversion level of three would gain (or be willing to pay; certainty equivalent) based on utilizing the forecasts compared to the historical mean. Ind are Fama-French industry-sorted portfolio returns. Monthly data is used. Data is obtained from Ken French's website. Recursive forecasting is performed with the estimation sample starting at January 1930 (anchored starting point) to December 1979. Forecast sample starts at January 1980 and runs to December 2010. PLS regression estimates obtained from SIMPLS algorithm of de Jong (1993) are used to forecast. Excess log returns are used, the CRSP value-weighted market index is used for market returns, and 3-month US Treasury bills are used for the risk-free rate.

Table 3.9: Market return forecast results

	6 SV	25 SV	100 SV	5 ind	10 ind	12 ind	17 ind	30 ind	49 ind
OOS R-squared	1.01%	1.10%	1.17%	1.32%	1.21%	1.27%	1.04%	1.04%	1.07%
OOS p-value	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
IS R-squared	1.12%	1.07%	1.30%	1.22%	1.23%	1.26%	1.12%	1.21%	1.26%
Utility gain	3.62%	3.45%	3.72%	2.84%	2.76%	2.82%	2.43%	2.35%	2.55%

This table shows the R-squared values from forecasting the market return. SV are the size-value sorted Fama-French portfolios. Ind are the various Fama-French industry-sorted portfolios. Out-of-sample (OOS) R-squared is based on the historical mean of the market return. OOS p-value is the p-value obtained by regressing the Clark and West (2007) adjusted-MSPE statistic on a constant, as in Li et al. (2013). IS R-squared is the percent variance of Y (market returns) explained by the first X-component. Utility gain is the SV are Fama-French size-value sorted portfolio returns. Ind are the various Fama-French industry portfolio returns. Monthly data is used. Data is obtained from Ken French's website. Recursive forecasting is performed with the estimation sample starting at January 1930 (anchored starting point) to December 1979. Forecast sample starts at January 1980 and runs to December 2010. PLS regression estimates obtained from SIMPLS algorithm of de Jong (1993) are used to forecast. Excess log returns are used, the CRSP value-weighted market index is used for market returns, and 3-month US Treasury bills are used for the risk-free rate. Utility gain is the annualized average return that a mean-variance investor with a risk aversion level of three would gain (or be willing to pay; certainty equivalent) based on utilizing the forecasts compared to the historical mean.

Figure 3.1:
PLS weights on PC factors formed from 25 size-value portfolios



This figure shows the PLS weights when PCA factors are used as the predictor variables when forecasting. Here, 25 PCA factors (PC above) are formed from the 25 size-value Fama-French portfolios. The weights are obtained from the PLS forecast regression results where excess market returns are the dependent variable. Recursive forecasting is performed with the estimation sample starting at January 1930 (anchored starting point) to December 1979. Forecast sample starts at January 1980 and runs to December 2010. PLS regression estimates obtained from the SIMPLS algorithm of de Jong (1993) are used to forecast. Excess log returns are used, the CRSP value-weighted market index is used for market returns, and 3-month US Treasury bills are used for the risk-free rate.

Chapter Four: *Arbitrage Pricing Restrictions and the Predictability of Stock Returns by Statistical Factor Analysis*

4.1. Introduction

The Arbitrage Pricing Theory (APT) of Ross (1976) provides the crucial insight that any set of factors covering the systematic risk of a particular group of assets should also explain the mean returns of these assets. Empirical implementation may be subdivided into three approaches. First, the macroeconomic sources approach as employed by Chen, Roll, and Ross (1986) in which a set of macro variables, thought on a priori theoretical grounds to contain all systematic risks, is used as the factors to explain asset returns. Second, the mimicking factors approach, as prominently employed by Fama and French (1993) and Fama and French (1996), for instance, in which a predetermined number of well-diversified portfolios of the assets to be evaluated are used as the risk factors.[29] The idea is that, as long as the factor portfolios are well diversified, sufficiently diverse, and equal in number to the true number of underlying factors, they will be linear combinations of the underlying factors and hence capture all systematic risk.

The third approach, the factor analysis approach by which factors are obtained statistically as linear combinations of the assets to be evaluated, is the focus of our paper.[30] We adapt the factor analytical approach here by including

[29] See also Hugerman, Kandel, and Stambaugh (1987) and Grinblatt and Titman (1987) for more on the mimicking portfolio approach.

[30] While there have been limited empirical evaluations of the APT recently, Lehmann and Modest (1988) provide a good overview of the empirical issues and estimation techniques of the APT.

the theoretical APT restrictions on mean returns to improve estimation and forecast performance under the maintained hypothesis that the APT is (approximately) correct. Roll and Ross (1980) and others utilize factor analysis (FA) or principal components analysis (PCA) to identify the portfolios that best explain the variation in realized returns of all assets under consideration. The number of such portfolios that have significant explanatory power for return covariances may be determined statistically from Bai and Ng (2002). Having established the factors that best explain covariances, subsequently the implication of the APT that exposures to this set of factors fully explain the differences in mean returns across assets may be tested.

Whereas FA focuses on finding the factors that explain systematic covariances and PCA focuses on finding the factors that explain total return covariances, Chamberlain and Rothschild (1983) show that asymptotically either approach will correctly identify the systematic risk factors because as the number of assets goes to infinity the systematic risk dominates any non-pervasive idiosyncratic risk. PCA allows the idiosyncratic risk matrix not to be diagonal so that returns may be correlated after correction for systematic risk as long as the idiosyncratic risk is not pervasive. PCA is also easy to implement and so we focus on (adapting) this approach.

Grinblatt and Titman (1987) introduce the idea of local efficiency and show that the APT implies that the factors are locally efficient. This also helps the APT avoid the Roll (1977) critique.

Based on the third approach, principal component factors are used as systematic asset pricing factors to test the APT and can then be applied in practice if the APT is not rejected. However, for purposes of obtaining factors that are useful for generating improved decisions in practical applications, dealing e.g. with costs of capital[31], hedging strategies, forecasting asset returns, evaluation of portfolio managers, and event studies, it is advantageous to impose the APT restrictions in advance as part of the factor estimation so that the resulting factor weights are more reliably estimated.

The objective in this paper is to provide a new method for obtaining statistical asset pricing factors that imposes a condition limiting the pricing errors for mean returns jointly with maximizing the variance in returns explained by the factors. In the language of the standard two-pass asset pricing approach, it maximizes the first-pass fit averaged across all test assets subject to a guaranteed minimum fraction of second-pass mean pricing errors. A related work is that of Nardari and Scruggs (2007), who impose the APT pricing restriction (that of mean returns being linear combinations of systematic factors) into a latent factor model with a multivariate stochastic volatility (MSV) process under a Bayesian

[31] For examples of asset pricing models used for cost of capital estimates, see Pratt and Grabowski (2008), Myers and Turnbull (1977), Goldenberg and Robin (1991), Bartholdy and Peare (2005), Fuller and Kerr (1981), and Cummins and Phillips (2005). Bruner et al. (1998), among others, provide survey evidence that practitioners typically use the CAPM in cost of capital estimation. Koedijk and van Dijk (2004) examine cost of capital from a global perspective with the CAPM, finding that the domestic market factor is sufficient for a firm to estimate cost of capital and that the CAPM and international CAPM provide little difference. Aside from estimates implied by asset pricing theory, previous research has used implied cost of capital measures (Li, Ng, and Swaminathan, 2013; Lee, Ng, and Swaminathan, 2011; Pástor, Sinha, and Swaminathan, 2008; Hou, Van Dijk, and Zhang, 2012; Easton, 2004) and accounting measures (Easton and Monahan, 2005) to estimate cost of capital.

methodology. They find their best model in terms of its Bayes factor is the model with three latent factors with MSV and the APT restrictions. Nardari and Scruggs use a Bayesian methodology, whereas we use an approach similar to PCA (PCA while imposing APT restrictions). They assign a unit weight to certain test assets (based on correlation with a preliminary first principal component) to form their three factors while we use linear combinations of all test assets to form optimal APT factors. This allows for an investor/practitioner to find the portfolio weights and form our factors in real time. We argue that our approach provides true APT factors (rather than selecting one portfolio to serve as a factor): a linear combination of all assets that is designed to capture covariance risk while perfectly explaining mean returns. Another difference is that we look out of sample whereas Nardari and Scruggs only look at in-sample posterior distributions and cross-sectional averages. Other differences include the following: Nardari and Scruggs allow for mean errors to be nonzero, whereas our main focus is on restricting mean errors to be zero; we do not make any assumptions on volatility; we use industry portfolios with no strong factor structure while they use ten size-sorted portfolios.

Another related work is MacKinlay and Pástor (2000), which assumes that a factor is missing from a linear factor model. They show that, under the strong assumptions of homoscedastic errors and uncorrelated errors, accounting for an unobserved factor amounts to the alpha showing up in the covariance matrix. Linking means and covariances allows for improved estimation of expected returns/means and higher Sharpe ratios in out-of-sample portfolio selection/optimization. They point out that they essentially are combining two factor approaches: FA/PCA, which focuses on covariance only, and factor mimicking portfolios (Fama-French three-factor model), which focuses on means only.[32] Our approach does the same: we link means and covariances by explaining mean returns up to an allowable error while maximizing covariance simultaneously. The benefits of our approach are as follows: we do not make any strong assumptions on the error terms as does MacKinlay and Pástor, and their approach requires conditioning on the Sharpe ratio of the missing factor (which must be pre-specified by the practitioner/econometrician). Since we impose that the number of specified factors perfectly price all assets, the latter issue is avoided by our approach. We also compare to the traditional asset pricing models (means only and covariances only as mentioned previously), which MacKinlay and Pástor do not. Finally, while MacKinlay and Pástor examine maximum possible Sharpe ratios out-of-sample, we implement an out-of-sample hedging approach.

While we perform forecasting of a hedged position rather than the raw portfolio returns with our derived factors, Simin (2008) examines the out-of-sample forecast performance of the CAPM and Fama-French three-factor model, finding that they perform poorly in predicting future asset returns compared to historical averages or constants. Simin further points out that statistical APT

[32] Pukthuanthong and Roll (2014) provide a way of identifying whether a factor is related to mean returns or risks (or both).

factors and macroeconomic factors perform worse than either of these two models. This differs starkly from our results, where the APT factors perform better than the CAPM and Fama-French three-factor model. One reason may be that Simin uses a rolling 60-month estimation window, while we find that a larger estimation window allows PCA factors to become more reliable (at least in terms of bias). Simin also uses the 25 size and book-to-market portfolios as the test assets, while we use industry portfolios. The Fama-French three-factor model can explain size and book-to-market portfolios well, but not industry portfolios.

One simple approach to performing PCA and also imposing zero pricing errors would be to restrict time series constants, “alphas”, to equal zero. However, in this case there is no requirement that mean pricing errors are zero so that an unknown level of pricing errors is allowed. Our approach, however, allows imposing in advance any desired mean return errors as a fraction of the squared sum of mean returns, including setting the fraction equal to zero.

An alternative way to view our method, if zero mean return errors are enforced (a 100% cross-sectional R-squared), is as an approach that generates a tangency portfolio as a portfolio of the factors and subject to this constraint finds the factors that also explain as much as possible of the variation in the time series of returns of all assets. Our approach differs from the Markowitz optimization approach, which has had limited success in out-of-sample application. For example, DeMiguel, Garlappi, and Uppal (2009) find that a naïve equal-weighted approach performs better out of sample in terms of its Sharpe ratio, turnover, and certainty equivalent return compared to the tangency portfolio. For an examination of investing in the tangency portfolio out of sample see also Best and Grauer (1991) and Michaud (1989). Kan and Zhou (2007) show the pitfalls of using sample estimates rather than population mean and covariance when performing out-of-sample mean-variance optimization, while also finding that a sample tangency portfolio and riskless asset can be beaten by adding the sample minimum-variance portfolio. This shows that using a larger sample/estimation window (as close to population as possible) may work better. Thus, we look at a large expanding estimation window (as well as a rolling window which does not perform as well perhaps due to this reason) when performing our out-of-sample analysis. Kan and Smith (2008) show that sample minimum-variance frontiers can be highly biased. Kan, Wang, and Zhou (2016) show that accounting for estimation error allows for portfolios that outperform the equal-weighted approach of DeMiguel, Garlappi, and Uppal (2009).

A fundamental issue in this paper is that an asset pricing model generates return forecasts. To evaluate the performance of these forecasts, we may simply calculate the mean square errors (MSE) generated from realized returns relative to the forecast. As we confirm below, PCA identifies factors that minimize these mean square errors, at least in sample. And this is true, irrespective of whether the APT holds. However, a typical decision maker employing the asset pricing model may distinguish between components of the errors that are systematic and caused by a mean bias in the asset pricing model, and errors that are simply part of return risk. From this perspective, it is important to consider separately the bias and the noise in the pricing model, as does Simin (2008). Simin (2008) and Leitch and Tanner (1991) both point out that investors/practitioners may have asymmetric loss functions, which requires a need to examine both bias and variance. Others that examine the limitations of symmetric loss functions/MSE and the implications of asymmetric loss functions include Patton and Timmermann (2007), Basu and Markov (2004), Clatworthy, Peel, and Pope (2012), and Elliot, Komunjer, and Timmermann (2008). Although they do not examine the breakdown of MSE, MacKinlay and Pástor (2000) do report both bias and MSE separately. Giacomini and White (2006) and Diebold and Mariano (2002) examine the issue of separating bias and variance (and the need for examining more than just MSE) from a forecasting perspective.

To provide an indication of the usefulness of imposing APT restrictions we apply our approach to a set of test assets and compare the results both in-sample and out-of-sample against those for other basic models. In particular, we consider the Fama-French 3-factor model as representative of a model typically applied and we consider a factor model obtained from conventional PCA. The test assets we consider are the 30 Industry portfolios obtained from Fama and French (available from Kenneth French's website). These assets constitute a good target as they do not have a strong factor structure and their mean returns have traditionally been difficult to explain (see Lewellen, Nagel, and Shanken, 2010; Fama and French, 1997; Chou, Ho, and Ko, 2012).

Imposing the APT restrictions perfectly will obviously result in a zero alpha/bias in sample, and PCA by design will have the lowest possible MSE in sample.[33] We do indeed find that the MSE when imposing the APT restrictions ("constrained" PCA or C-PCA) is only slightly higher while completely eliminating bias. To judge the out-of-sample performance we consider a hedging approach where the in-sample portfolio weights and betas are used to hedge each portfolio over the next month. While this is not a true forecasting approach in that factor realizations are used, it is nevertheless an approach that is applicable in real time. The out-of-sample results show that again the MSEs are not significantly different between PCA and C-PCA. However, C-PCA reduces the bias out of sample by around 30% compared to PCA.

We argue that any investor or practitioner who has a long-term investment horizon in mind would prefer C-PCA due to the reduced bias, regardless of loss function.[34] In our out-of-sample application we re-estimate and change the hedged position each month, and then evaluate performance over the full out-of-sample window. While it is possible that an investor would still prefer reduced

[33] Comparing the in-sample cross-sectional R-squared of the models does not provide much insight either; Kan, Robotti, and Shanken (2013) investigate the significance of differences between cross-sectional R-squared values, rather than a simple comparison of the values, and they find that most models are not significantly different.

[34] Investment horizon also has a clear monotonic relationship with bias, whereas various loss functions may have contrasting implications. Also, asymmetric loss functions may place more weight on bias *or* variance depending on the specification/investor or analyst.

bias over this large sample period/window, it is also conceivable that the investor would implement a longer forecast horizon and avoid monthly re-estimation and rebalancing. As the forecast horizon lengthens, the target return will approach the mean return, so C-PCA should perform better regardless of loss function (its MSE should improve relative to all other models). This is also consistent with the findings of MacKinlay and Pástor (2000): as the sample size gets larger (240 and/or 360 months) the unbiased/unrestricted estimator they use performs better than their restricted estimator that links means and covariances but is biased. We also provide a theoretical and empirical discussion of the relationship between investment horizon and bias: lower bias at a smaller horizon is preferable as the horizon grows. Thus, lower bias will mean as the horizon is increased, the particular model will perform better in terms of MSE compared to a model with a higher bias.

Higher MSE due to higher variance is akin to higher month-to-month fluctuations around the target (albeit only slightly higher MSE/variance for C-PCA), while lower bias means being closer to the target on average. For example, for a long-term project, a firm would want a cost of capital estimate that is as accurate as possible, while the firm may not care much about short-term variations. If the application emphasizes bias (average accuracy) even slightly in comparison to variance, then this particular practitioner prefers C-PCA. C-PCA also has a better second pass fit both in sample and out. We additionally find that

PCA and C-PCA are both significantly better than the traditional asset pricing models for these industry portfolios.

While the application of the derived C-PCA pricing factors that we focus on is out-of-sample hedging, there are many potential applications of these factors. There has been extensive research applying APT/PCA factors to various settings.[35] Kozak, Nagel, and Santosh (2014) find that a stochastic discount factor based on PCA can explain away many anomalies out-of-sample. They also show it may not be possible to distinguish whether these factors are sentiment-based or risk-based. Bussière, Hoerova, and Klaus (2015) apply PCA to hedge fund returns in order to construct the main factor for a set of hedge fund assets and examine changes in the hedge funds' sensitivity to the factor over time. Herskovic et al. (2016) develop a common idiosyncratic volatility (CIV) factor, which they note can be seen as an APT factor. They find this factor to be priced and to help explain various pricing anomalies. One of the models used to find CIV (residual variance) is the first five principal components. Ludvigson and Ng (2009) conduct factor analysis on a panel of macroeconomic variables to construct macro factors which are used to forecast bond returns, and examine bond risk premia. Goyal, Pérignon, and Villa (2008) find that there are two common APT factors in both the NASDAQ and NYSE, but each one has another separate risk factor that does not extend to the other market. Chen, Hsieh, and

[35] Many implementations of PCA, especially those for forecasting purposes, follow either the methodology of Stock and Watson (2002) or Forni et al. (2000). Stock and Watson use static principal components while Forni et al. use dynamic principal components, although both use dynamic factor models to forecast.

Jordan (1997) applies the APT to the real estate market, by examining both macroeconomic risk factors and derived factors as they relate to REIT returns. They find that for some time periods the macro factors explain REIT returns better, while in other periods the two models perform equally. An example of an application of macroeconomic factors is Kavussanos, Marcoulis, and Arkoulis (2002), which applies macroeconomic factors to international industry returns and examine the investment implications of their findings.

In the following we first review in Section 4.2 the basic theory regarding the PCA approach and its optimality properties. In Section 4.3 we present the derivation of an approach that optimally imposes the APT pricing restrictions when extracting factors from a group of assets based on applying the Ky Fan maximum principle for Rayleigh quotient pencils. Section 4.4 applies the approach to explain the returns of the thirty Fama-French industry portfolios in and out of sample and contrasts the performance with alternative approaches. Section 4.5 concludes.

4.2. Basic Properties of Statistical Factor Analysis

4.2.1 Returns Specification with Factor Structure

Consider the returns of a group of assets as given by

$$(4.1) \qquad \mathbf{r} = \boldsymbol{\alpha} + \mathbf{B}\mathbf{Q}'\mathbf{r} + \mathbf{e},$$

where $\mathbf{r}, \boldsymbol{\alpha}, \mathbf{e}$ are *n x 1* vectors indicating, respectively, the set of returns, the mean pricing errors, and the non-systematic pricing errors for each of *n* risky assets. The matrix $\mathbf{Q}$ is an *n x k* matrix of full column rank that represents for each of the *k* factors the weights put on the *n* assets. Thus, $\mathbf{r_k} = \mathbf{Q}'\mathbf{r}$ represents the *k x 1* vector of factor returns. The factor loadings (betas) for all assets are given by the *n x k* matrix $\mathbf{B}$. We define the *n x n* covariance matrix of returns as $\boldsymbol{\Sigma} = \mathbf{E[(r - \boldsymbol{\mu})(r - \boldsymbol{\mu})']}$. The distribution of returns is constant over time, and, by extension, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are time invariant. Note that any set of returns may be represented by equation (4.1) without loss of generality such that $E(\mathbf{e}) = 0$ and such that $E(\mathbf{r_k e}') = \mathbf{0}$.

Taking expectation in equation (4.1), the mean returns of all assets are given as

(4.2) $\boldsymbol{\mu} = \boldsymbol{\alpha} + \mathbf{BQ}'\boldsymbol{\mu}$.

4.2.2 Factors Chosen to Minimize Errors and PCA

We show first that selecting factors to minimize the sum of squared non-systematic pricing errors is equivalent to adopting the principal components as factors. While this result was noted previously (Jong and Kotz, 1999), it is not well known in the finance literature. We emphasize that PCA does not simply select the linear combinations of assets that best captures the communalities in returns, but that adopting these components as factors provides the lowest attainable quadratic variation in the return errors:

$$\underset{\mathbf{B,Q}}{Min} \quad \left[E(\mathbf{e}'\mathbf{e})\right].$$

Since $E(\mathbf{e}'\mathbf{e}) = Tr[E(\mathbf{ee}')]$, where Tr represents the trace of the matrix, minimizing the sum of squared errors from equations (4.1) and (4.2) amounts to:

$$(4.3) \qquad \underset{\mathbf{B,Q}}{Min} \quad Tr\left[(\mathbf{I}-\mathbf{BQ}')\,\mathbf{\Sigma}\,(\mathbf{I}-\mathbf{QB}')\right].$$

First, for given factor weights, choose the factor loadings **B** that minimize the squared errors. Matrix differentiation in equation (4.3) and finding the stationary point implies $\mathbf{\Sigma Q}-\mathbf{B}(\mathbf{Q}'\mathbf{\Sigma Q})=0$ so that

$$(4.4) \qquad \mathbf{B}=\mathbf{\Sigma Q}(\mathbf{Q}'\mathbf{\Sigma Q})^{-1},$$

the standard formula for obtaining multivariate betas. Plug equation (4.4) into equation (4.3) to find the concentrated objective:

$$(4.5) \qquad \underset{\mathbf{Q}}{Min} \quad Tr\left[\mathbf{\Sigma}-\mathbf{\Sigma Q}(\mathbf{Q}'\mathbf{\Sigma Q})^{-1}\mathbf{Q}'\mathbf{\Sigma}\right].$$

Since Tr ($\mathbf{\Sigma}$) is constant and using the property that Tr ($\mathbf{AB}$) = Tr ($\mathbf{BA}$) equation (4.5) becomes

$$(4.6) \qquad \underset{\mathbf{Q}}{Max} \quad Tr\left[(\mathbf{Q}'\mathbf{\Sigma}^2\mathbf{Q})(\mathbf{Q}'\mathbf{\Sigma Q})^{-1}\right].$$

Taking the "square root" of the positive definite $\mathbf{\Sigma}$ and defining $\mathbf{Q}^* = \mathbf{\Sigma}^{½}\mathbf{Q}$ we obtain

$$\text{(4.7)} \qquad \underset{\mathbf{Q}^*}{Max} \quad Tr\left[(\mathbf{Q}^{*}{'}\mathbf{\Sigma}\mathbf{Q}^*)(\mathbf{Q}^{*}{'}\mathbf{Q}^*)^{-1}\right].$$

The trace in equation (4.7) is known as the Rayleigh Quotient which by the Ky Fan maximum principle (Fan, 1949. See also for instance Bhatia, 2013, p. 35, and Li, 2015) is maximized when the *n x k* matrix $\mathbf{Q}^*$ is orthonormal and consists of the *k* eigenvectors associated with the *k* largest eigenvalues of matrix $\mathbf{\Sigma}$ (and is minimized if $\mathbf{Q}^*$ contains the *k* eigenvectors corresponding to the *k* smallest eigenvalues of $\mathbf{\Sigma}$). Thus, we know that

$$\text{(4.8)} \qquad \mathbf{\Sigma}\mathbf{Q}^* = \mathbf{Q}^*\mathbf{\Delta}_k \text{, as well as } \mathbf{\Sigma}\mathbf{Q} = \mathbf{Q}\mathbf{\Delta}_k$$

where $\mathbf{\Delta}_k$ is the *k x k* diagonal matrix containing the *k* largest eigenvalues of $\mathbf{\Sigma}$. Pre-multiplying both sides of equation (4.8) by $\mathbf{Q}^{*}{'}$ (or the second equation by $\mathbf{Q}'$) shows that the maximum trace of the Rayleigh Quotient in equation (4.7) is equal to $Tr\,(\mathbf{\Delta}_k)$ which is the sum of the *k* largest eigenvalues. Also, we find that $Tr(\mathbf{Q}'\mathbf{\Sigma}^2\mathbf{Q})(\mathbf{Q}'\mathbf{\Sigma}\mathbf{Q})^{-1} = Tr(\mathbf{Q}'\mathbf{\Sigma}\mathbf{Q})(\mathbf{Q}'\mathbf{Q})^{-1}$. It follows that either the eigenvectors $\mathbf{Q}^*$ or $\mathbf{Q}$ may be used to generate the *k* factors $\mathbf{r}_k^* = \mathbf{Q}^{*}{'}\mathbf{r}$ or $\mathbf{r}^k = \mathbf{Q}'\mathbf{r}$ that minimize the unexplained variation. Note that equation (4.8) satisfies the first-order conditions for a stationary point in $\mathbf{Q}^*$ or $\mathbf{Q}$ subject to a normalization constraint $\mathbf{Q}'\mathbf{\Sigma Q} = \mathbf{I}_k$: if we define the Lagrangian $L = Max\, Tr[(\mathbf{Q}'\mathbf{\Sigma}^2\mathbf{Q}) - \Delta_k(\mathbf{Q}'\mathbf{\Sigma Q} - \mathbf{I}_k)]$ it generates equation (4.8). The second-order conditions do not hold, however, which is the reason that we need to appeal to the properties of Rayleigh Quotients to generate the result. We can view the fraction of return variation explained by the *k* factors as $Tr\ (\mathbf{\Delta}_k)\ /\ Tr\ (\mathbf{\Sigma})$, the sum of the *k* largest eigenvalues divided by the sum of all eigenvalues. The $\mathbf{Q}^*$ or $\mathbf{Q}$ matrix obtained via equation (4.8) consists of the first *k* principal components and, by Ky Fan's theorem, the optimal realized objective of equation (4.6) equals the sum of the *k* largest eigenvalues of the $\mathbf{\Sigma}$ matrix.

4.3. Incorporating Arbitrage Pricing Restriction

4.3.1 Adding APT Constraint

If the factor returns explain most of the variance in returns so that the remaining variance is non-pervasive and may be diversified (Chamberlain and Rothschild, 1983), then the APT of Ross (1976) implies that the $\boldsymbol{\alpha}$'s are (approximately) zero. To optimally generate the statistical set of factors characterized by $\mathbf{Q}$ and obtain factor returns $\mathbf{r}_k = \mathbf{Q}'\mathbf{r}$ we now consider the following optimization problem:

$$\text{(4.9)} \qquad \underset{\mathbf{B},\mathbf{Q},\lambda}{Min} \quad \left[E(\mathbf{e}'\mathbf{e}) + \lambda(\boldsymbol{\alpha}'\boldsymbol{\alpha} - \gamma\,\boldsymbol{\mu}'\boldsymbol{\mu})\right],$$

where $0 \le \gamma < 1$ represents the target level for the sum of squared alphas relative to the sum of squared mean returns. Thus, γ represents the targeted fraction of mean return errors allowed, i.e., the sum of squared mean pricing errors compared to the sum of squared means. Most often we will consider $\gamma = 0$ as the target which means that all mean pricing errors, the "alphas", are constrained to be zero.

Using equations (4.1) and (4.2), the properties of traces, and taking expectations we may replace equation (4.9) by

$$(4.10) \quad \underset{\mathbf{B},\mathbf{Q},\lambda}{Min} \quad Tr\left[(\mathbf{I}-\mathbf{BQ}')\,\mathbf{V}(\lambda)\,(\mathbf{I}-\mathbf{QB}')\right]-\lambda\gamma\,\boldsymbol{\mu}'\boldsymbol{\mu},$$

$$\mathbf{V}(\lambda) \equiv \boldsymbol{\Sigma} + \lambda\,\boldsymbol{\mu}\boldsymbol{\mu}'.$$

Analogously to the unconstrained case, we have

$$(4.11) \quad \mathbf{B} = \mathbf{V}(\lambda)\mathbf{Q}[\mathbf{Q}'\mathbf{V}(\lambda)\mathbf{Q}]^{-1},$$

and in addition the constraint implies that $\boldsymbol{\alpha}'\boldsymbol{\alpha} = \gamma\,\boldsymbol{\mu}'\boldsymbol{\mu}$. Note that the objective in equation (4.9) requires that the optimally estimated betas, $\mathbf{B}$, are not the standard OLS betas of equation (4.4). The relevant association between asset return and factors here also involves consideration of the mean returns. The concentrated objective becomes after eliminating $\mathbf{B}$:

$$(4.12) \quad \underset{\mathbf{Q}}{Max} \quad Tr\left([\mathbf{Q}'\mathbf{V}(\lambda)^2\mathbf{Q}][\mathbf{Q}'\mathbf{V}(\lambda)\mathbf{Q}]^{-1}\right),$$

where λ is set to force $\boldsymbol{\alpha}'\boldsymbol{\alpha} = \gamma\,\boldsymbol{\mu}'\boldsymbol{\mu}$. Similar to the unrestricted case the eigenvectors are found from

$$\text{(4.13)} \quad \mathbf{V}(\lambda)\mathbf{Q} = (\boldsymbol{\Sigma} + \lambda\,\boldsymbol{\mu}\boldsymbol{\mu}')\mathbf{Q} = \mathbf{Q}\boldsymbol{\Delta}_k .$$

Pre-multiplying by $\mathbf{Q}'$ and subsequently pre- and post-multiplying both sides by $\boldsymbol{\mu}'\mathbf{Q}$ and its transpose, respectively, and using the normalization that $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_k$ implies for the optimal choice of $\mathbf{Q}$ that $\boldsymbol{\mu}_k'(\boldsymbol{\Delta}_k - \mathbf{Q}'\boldsymbol{\Sigma}\mathbf{Q})\boldsymbol{\mu}_k = \lambda(\boldsymbol{\mu}'\mathbf{Q}\mathbf{Q}'\boldsymbol{\mu})^2$. To impose $\boldsymbol{\alpha}'\boldsymbol{\alpha} = \gamma\,\boldsymbol{\mu}'\boldsymbol{\mu}$ eliminate $\boldsymbol{\alpha}$ from equation (4.2). Equations (4.11) and (4.13) imply $\boldsymbol{\mu}'\mathbf{Q}\mathbf{Q}'\boldsymbol{\mu} = (1-\gamma)\,\boldsymbol{\mu}'\boldsymbol{\mu}$. Thus, imposing $\boldsymbol{\alpha}'\boldsymbol{\alpha} = \gamma\,\boldsymbol{\mu}'\boldsymbol{\mu}$ for optimal $\mathbf{Q}$ into $\boldsymbol{\mu}_k'(\boldsymbol{\Delta}_k - \mathbf{Q}'\boldsymbol{\Sigma}\mathbf{Q})\boldsymbol{\mu}_k = \lambda(\boldsymbol{\mu}'\mathbf{Q}\mathbf{Q}'\boldsymbol{\mu})^2$ produces

$$\text{(4.14)} \quad \lambda = \boldsymbol{\mu}_k'(\boldsymbol{\Delta}_k - \mathbf{Q}'\boldsymbol{\Sigma}\mathbf{Q})\boldsymbol{\mu}_k /(1-\gamma)^2(\boldsymbol{\mu}'\boldsymbol{\mu})^2 .$$

Solving equations (4.13) and (4.14) jointly generates both the optimal $\mathbf{Q}$ and λ. Accordingly, the *k* factor returns are $\mathbf{r}^k = \mathbf{Q}'\mathbf{r}$ which guarantee the lowest non-systematic errors subject to the mean pricing errors corresponding to a target fraction $\gamma = \boldsymbol{\alpha}'\boldsymbol{\alpha} / \boldsymbol{\mu}'\boldsymbol{\mu} \geq 0$.

The Appendix provides a Matlab program for solving equations (4.13) and (4.14) simultaneously. The Lagrangian multiplier λ must be determined so that the optimization weight on reducing the alphas is sufficient to deviate enough from minimizing overall pricing errors to meet the alpha target. The resulting optimal $\mathbf{Q}$ must generate a mean square error in-sample that is higher than in the standard PCA case, but produces a larger cross-sectional R-squared that would rise to 100% for $\gamma = 0$. The approach comes down to calculating the eigenvalues and eigenvectors of matrix $\mathbf{V}(\lambda) = \mathbf{\Sigma} + \lambda\boldsymbol{\mu}\boldsymbol{\mu}'$ and using the eigenvectors of the k largest eigenvalues as follows from the Ky Fan maximum principle. As in standard PCA the factor loadings may be obtained from $\mathbf{B} = \mathbf{Q}$ so that for the case of $\gamma = 0$ (zero alphas) from equation (4.1) the predicted returns equal $\hat{\mathbf{r}} = \mathbf{Q}\mathbf{Q}'\mathbf{r} = \mathbf{Q}\mathbf{r}_{\mathbf{k}}$.

4.3.2 Imposing OLS Betas

The standard approach for determining the unexplained component of asset returns implied by a particular asset pricing model is to calculate alphas from "first-pass" time series regressions of the asset returns on the factor returns. In this approach the factor sensitivity estimates are usually obtained as OLS betas: The empirical results for the (non-optimal) case of OLS betas, $\mathbf{B} = \mathbf{\Sigma Q}(\mathbf{Q}'\mathbf{\Sigma Q})^{-1}$, when the factors are specified as $\mathbf{Q}'(\mathbf{r} - \boldsymbol{\mu})$ are available upon request from the authors. While this does not provide the minimum MSE it is a reasonable case to consider as it corresponds to imposing a constraint on alphas as typically calculated. For reasons of tractability we focus here on the weighted alphas which relate to the difference in maximum squared Sharpe ratios of the assets compared to the factors as discussed by Gibbons, Ross, and Shanken (1989): $\boldsymbol{\alpha}'\mathbf{\Sigma}^{-1}\boldsymbol{\alpha} = \boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}'\mathbf{Q}(\mathbf{Q}'\mathbf{\Sigma Q})^{-1}\mathbf{Q}'\boldsymbol{\mu}$. We set the weighted alphas equal to a target fraction of the weighted mean returns: $\boldsymbol{\alpha}'\mathbf{\Sigma}^{-1}\boldsymbol{\alpha} = \gamma(\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu})$. As before, if we set $\gamma = 0$ then this implies the constraint $\boldsymbol{\alpha} = 0$ (for all assets). The Appendix derives that for the case of

$$(4.15)\quad \underset{\mathbf{Q},\lambda}{Min} \quad \left[E(\mathbf{e}'\mathbf{e}) + \lambda(\boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha} - \boldsymbol{\gamma}\,\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})\right],$$

subject to equations (4.1) and (4.2) and $\mathbf{B} = \boldsymbol{\Sigma}\mathbf{Q}(\mathbf{Q}'\boldsymbol{\Sigma}\mathbf{Q})^{-1}$, that the factors are found first from

$$(4.16)\quad \mathbf{V}(\lambda)\,\mathbf{Q} = (\boldsymbol{\Sigma} + \lambda\,\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-\frac{1}{2}})\mathbf{Q} = \mathbf{Q}\,\boldsymbol{\Delta}_k\,.$$

using the normalization that $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_k$ and, second defining $\boldsymbol{\mu}_{\mathbf{k}} = \mathbf{Q}'\boldsymbol{\mu}$, from

$$(4.17)\quad \lambda = \boldsymbol{\mu}_k'(\,\boldsymbol{\Delta}_k - \mathbf{Q}'\boldsymbol{\Sigma}\mathbf{Q})\boldsymbol{\mu}_k\,/(1-\gamma)^2\,(\,\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^2\,.$$

Then, subject to a target fit in the second pass, the factors maximize the first-pass fit.

4.3.3 Sub-Optimal Direct Restriction on the Alphas

An alternative way of reducing the mean pricing errors, as mentioned previously, is to omit the constant in estimating $\mathbf{B}$ and $\mathbf{Q}$. Thus, in equation (4.1) we set $\boldsymbol{\alpha}$ **=0.** Then $E(\mathbf{e}'\mathbf{e})$ is given as $Tr\left[(\mathbf{I} - \mathbf{B}\mathbf{Q}')\,E(\mathbf{r}\mathbf{r}')\,(\mathbf{I} - \mathbf{Q}\mathbf{B}')\right]$. Since $E(\mathbf{r}\mathbf{r}') = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}'$, the estimation problem becomes similar to both the restricted and unrestricted cases but with $\lambda = 1$ instead of the constrained on the mean pricing errors. Due to the omission of the constant, $E(\mathbf{e}) \neq 0$ so that mean pricing errors are not avoided, as is the case in Nardari and Scruggs (2007). In fact, the fraction of squared mean pricing errors may be obtained from the constrained PCA case by searching for the particular target fraction of mean pricing errors, γ, that implies $\lambda = 1$.

4.3.4 Correlation Matrix instead of Covariance Matrix

PCA analysis is often based on a correlation matrix rather than a covariance matrix. See, for instance, Jolliffe (2002). While working with returns implies there is no serious need for further normalization, there may still be cases in which a correlation matrix is preferable.

In effect, we now examine normalized returns for which $\mathbf{r}^* = \mathbf{S}^{-1}\mathbf{r}$, where **S** is an *n x n* diagonal matrix containing the *standard deviations* of the returns on the diagonal such that $\mathbf{S}^2 = \mathbf{\Sigma}$. The covariance matrix of $\mathbf{r}^*$ is then the correlation matrix of $\mathbf{r}$. The only difference compared to the covariance analysis case is in the interpretation of the results. In particular, maximizing explained covariance (first-pass) now becomes maximizing explained correlation, which is in turn equivalent to maximizing the squared correlation. The latter amounts to maximizing the equal-weighted average of the R-squares for the *n* first-pass regressions (subject to a restriction on the squared mean pricing errors for the normalized returns).

4.4. Application to Industry Portfolios

4.4.1 Number of Factors and MSE Breakdown

The in-sample and out-of-sample performances of PCA and C-PCA will be compared, along with the traditional asset pricing models (CAPM, Fama-French three-factor model (FF3), and Carhart, from Carhart, 1997), using the thirty Fama-French industry-sorted monthly portfolios. The industry portfolios are chosen due to their weak factor structure. The sample starts in January 1927, when the momentum factor is first available. The sample ends in December 2015. Excess returns are used (with three-month T-bill used as the risk-free rate), so that each portfolio can be considered as a zero-investment portfolio.

First, we investigate the optimal number of factors to implement. While there have been many that investigate the appropriate number of factors in PCA/FA (see Connor and Korajczyk, 1993), a widely used test is that of Bai and Ng (2002).[36] Bai and Ng develop a test for factor models with large panels. The loss functions they implement penalize in both the time and cross-sectional dimension. They propose three variations of a panel information criterion (IC), which do not depend on the specified maximum number of factors, and three variations of a panel C_p[37] criterion (PC) which implement the Mallows (1973) C_p criterion in a panel setting. Table 4.1 shows that five of the six criteria

[36] See also Harvey, Liu, and Zhu (2015) for an examination of the statistical hurdle that a new factor must clear in order to be accepted.

[37] The C_p criterion stands for the Mallows (1973) information criterion for a model with P regressors.

suggest k=4 factors for the set of industry portfolios, with the last PC criterion suggesting k=5 factors. Therefore, we will use four factors as our base model (with PCA and C-PCA both having four factors unless otherwise specified).

[Table 4.1 goes here]

Before examining the results, the breakdown of MSE should be discussed further. MSE in this context can be broken down into two parts: the variance of the error term and the bias squared (average error squared). The simple summation of the two gives MSE; this implies putting an equal weight on each, however. We argue there are situations such as those related to cost of capital, hedging, long-term investments/projects, etc. where an investor would care more about bias. This equates to caring more about on average how far from the target (bias) an investor/practitioner/firm is rather than about the short-term fluctuations from the target (variance). In finance and forecasting especially, most researchers report MSE but do not consider the two components separately. The breakdown is as follows (using the definition of variance):

$$(4.18) \quad MSE \equiv \frac{1}{T}\sum_{t=1}^{T}\left[(r_t - \hat{r}_t)^2\right] \equiv \left[\frac{1}{T}\sum_{t=1}^{T}(r_t - \hat{r}_t)\right]^2 + \sigma_\varepsilon^2 \ ,$$

where r is the realized return, $\hat{r}$ is the predicted return, and σ_ε^2 is the variance of the error term $\varepsilon = (r - \hat{r})$. The first term on the right-hand side in (18) is the bias squared, and the last term is the variance term. We will mostly focus on the absolute level of the bias, the square root of the bias squared (in order to give an economic interpretation), and the MSE.

4.4.2 Expanding Window

Panel A of Table 4.2 shows the in-sample results of PCA, C-PCA, the CAPM, the FF3 model, and the Carhart model. Reported are the MSE (multiplied by 10,000) and absolute value of the bias multiplied by 100 to be in percent form (bias is the mean of the thirty portfolios is given for each test statistic), as well as the second pass adjusted R-squared (using mean returns). Note that this implementation of C-PCA does not use OLS alphas; therefore, the 2nd pass fit will not be 100%. We know two things will be true of the in-sample results: first, PCA will have the lowest possible MSE. Second, C-PCA with $\gamma = 0$ will have zero bias (with a much better second pass fit).[38] Kan, Robotti, and Shanken (2013) argue that simple comparisons of second-pass R-squared values are not sufficient or may induce misleading conclusions.[39] The MSEs of C-PCA and PCA are not significantly different, however. The interpretation is that an investor would only have to sacrifice a bit of MSE (due to variance) in order to achieve zero bias. Also, the panel shows that the traditional asset pricing models perform much worse in all areas.

[38] As discussed in Section 4.3.2, the alphas and betas we use in our main approach differ from the traditional OLS betas and alphas. Therefore, these C-PCA factors will not achieve a perfect second-pass fit, as they are designed to achieve zero error when applying the weights and betas (which are the same) directly from the approach, rather than using the weights as regressors in a second-pass regression. An alternative C-PCA methodology achieves zero OLS alphas and a perfect second-pass fit. The empirical results of this alternative approach are available from the authors upon request.

[39] Kan, Robotti, and Shanken (2013) examine R-squared differences when the models have beta misspecification. However, C-PCA would not apply here as a perfect fit is imposed.

[Table 4.2 goes here]

Since we know that PCA will have the lowest MSE and variance while C-PCA will have zero bias by construction, we now turn to the out-of-sample results for a more meaningful comparison. Panel B shows the results when out-of-sample estimation begins in January 1977, with an anchored starting point for the estimation window of January 1927 so that the estimation window begins at 50 years. Recursive monthly estimation is performed, with an expanding estimation window. The out-of-sample analysis here can be thought of as a hedging approach. The portfolio weights and betas (for PCA and C-PCA, the weights and betas are the same) are applied one month ahead to the portfolio realizations. Thus, the portfolio betas are applied to the factor realizations to provide a predicted return for each portfolio. While realizations are used (rather than a pure forecasting approach), this strategy is still implementable by an investor in real time. The portfolio weights can be found prior to the upcoming month via the in-sample estimation. The investor would then hold the portfolios with the given weights in order to form the factors for the next month. Then, the in-sample betas are used as factor weights. If the model holds perfectly out-of-sample, then each portfolio would be hedged against systematic risk perfectly out-of-sample.

Panel B shows the MSE and the absolute value of the bias for each model out of sample. Again, PCA has a slightly lower MSE compared to C-PCA, while the other models perform much worse. Whereas C-PCA obviously no longer has a zero bias, it is still much lower than PCA, with a roughly 30% reduction in bias. This translates to an average error of 1.8% for PCA annually, versus 1.3% for C-PCA. If an investor wished to eliminate the systematic risk of these industry portfolios for any given application, she would have a slightly higher MSE (and variance) with C-PCA. This would mean the predicted return has more fluctuation around the realized return for a given month compared to PCA. However, in the long run (from 1977 to 2015), the investor's average error is roughly 0.5% lower annually (which equates to 19% over the sample). We argue that for certain applications and for those more concerned with the long-run performance rather than month-to-month fluctuations, bias would be of more importance. In other words, imposing the APT restrictions allows an investor to be closer to the target *on average*. Those more concerned with less error on average rather than short-term fluctuations would prefer C-PCA to PCA.

Panel B of Table 4.2 also reports a second pass fit and an error ratio. In this instance, the second pass fit is the adjusted R-squared of a regression of mean returns on mean predicted returns for the thirty portfolios. This gives an idea how well the predicted returns explain returns on average. The C-PCA model again is superior to the competing models. The error ratio is the ratio of the sum of the squared errors (mean returns less mean predicted returns, squared) to squared mean returns.[40] Hence, a lower ratio means lower errors on average, and C-PCA has a ratio much lower than the other models.

[40] Note that this is similar to, but not exactly the same as, 1 – second pass fit.

Panel C shows the results when only an initial estimation window of 60 months is used, so that hedging starts in January 1932. This not only provides a larger sample, but also allows for comparison to the results that follow where a rolling 60-month window is used. Again, imposing the APT restrictions reduces bias greatly (around 33% on average). C-PCA provides a better fit and also now a slightly lower MSE compared to PCA. However, the overall fit based on second pass fit and error ratio compared to Panel B in this case is worse. Again, the competing models perform much worse compared to PCA and C-PCA.

4.4.3 Bias and Random Error

The MSE may be written as

$$(4.19)\quad MSE \equiv E(r_t - \hat{\mu}_t)^2 \equiv E[(r_t - \mu_t) + (\mu_t - \hat{\mu}_t)]^2,$$

where

$$(4.20)\quad r_t = \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim IID(0, \sigma_\varepsilon^2).$$

$$(4.21)\quad \hat{\mu}_t = \mu_t + \eta_t, \qquad \eta_t \sim IID(b, \sigma_\eta^2).$$

The bias is given by *b*. Since ε_t is uncorrelated with anything we can write equation (1) as

$$(4.22)\quad MSE = \sigma_\varepsilon^2 + \sigma_\eta^2 + b^2.$$

The PCA estimation of $\hat{\mu}_t$ provides the lowest possible *MSE based on the perspective of the observation frequency*. I.e., it effectively minimizes the *MSE* for the one-period ahead forecast since the observation frequency is (by definition) one period, a month for our data. We may, however, be interested in a

T-period forecast horizon. In that case, given the IID assumptions, the *MSE* becomes:

$$(4.23)\quad MSE(T) = T(\sigma_\varepsilon^2 + \sigma_\eta^2) + (Tb)^2 .$$

Thus, if we define the fraction of MSE due to bias as $\lambda \equiv b^2 / (\sigma_\varepsilon^2 + \sigma_\eta^2)$, then the *MSE* per unit of the forecast horizon is:

$$(4.24)\quad MSE(T)/T = (\sigma_\varepsilon^2 + \sigma_\eta^2)(1 + \lambda T) .$$

It is constant if the forecast is not biased and otherwise, whether the bias is positive or negative, increases linearly with the forecast horizon *T*.

Thus, if we compare two forecasts, based on PCA and C-PCA, such that:

$MSE_{PCA}(1) < MSE_{C-PCA}(1)$ and $\lambda_{PCA} > \lambda_{C-PCA}$ (both conditions should hold in our set-up), then there is a critical T=T^* such that for all $T<T^*$ we have $MSE_{PCA}(T) < MSE_{C-PCA}(T)$ and for all $T>T^*$ we have $MSE_{PCA}(T) > MSE_{C-PCA}(T)$.

4.4.4 Varying time horizon

While it is possible to have an asymmetric loss function that does not use MSE as the loss function and therefore prefer less bias to variance, the time horizon of the investment or forecast matters as well. Specifically, we argue (and show empirically) that as the out-of-sample time horizon increases, bias matters proportionately more compared to the variance term. This section performs the same out-of-sample risk hedging of the previous sections, but varies the out-of-sample time horizon. The results from a one-month horizon are shown again for comparison purposes, but now 3, 12, and 48-month horizons are added. The out-of-sample estimation starts after 60 months, so that an initial estimation window of 60 months is implemented, expanding to add the most recent month. Again, the sample is January 1927 to December 2015, with the out-of-sample window starting in January 1932 and ending in December 2015. Monthly returns are still used to find the loadings for risk hedging purposes. Table 4.3 shows the results for the C-PCA and PCA methods.

Table 4.3 provides the MSE, root MSE (RMSE), bias, bias squared, and variance and standard deviation of the error term. Bias, standard deviation, and RMSE are multiplied by 100 and reported in percent form. MSE, bias squared, and variance are multiplied by 10,000 to be consistent. The RMSE and standard deviation are provided in order to give units in percentage terms. Also, the proportion of bias to RMSE is provided as well in order to give an economic magnitude of how much bias contributes to the investor's loss (RMSE).[41] The first thing that can be seen is that extending the horizon for hedging purposes results in C-PCA performing relatively better compared to PCA in terms of MSE. At a one-month horizon, PCA has a MSE that is 0.51% higher than C-PCA. At three months this proportionate increase rises to 1.14%, and it is 3.01% higher at

[41] It should be noted that bias over RMSE and standard deviation over RMSE do not sum to one. The reason is that there are covariance terms that are ignored. However, the ratio of bias to RMSE does provide a better economic interpretation than bias squared over MSE, since the units (percent returns) are unchanged. It should also be pointed that while Equations 4.17 – 4.24 assume that the covariance terms are zero this is not the case empirically, as the covariance terms are noticeable.

twelve months. However, it does drop to 0.76% at 48 months, perhaps due to lost forecast power. Nonetheless, the less biased C-PCA factors perform relatively better than PCA as the time horizon increases, even when using MSE as the loss function.

[Table 4.3 goes here]

The results in Table 4.3 also show that the bias increase for both models is relatively linear and proportionate to the horizon. Going from one month to three months results in a bias that is very close to three times larger (and twelve times larger for twelve months). The variance terms are also close to linear increases as the horizon increases, albeit with slightly larger increases compared to bias. However, bias squared is what matters for MSE, so the increase from one month to three months is actually times nine, and times 144 for twelve months (assuming exactly linear increases, which breaks down a bit as the horizon approaches 12 and 48 months). Therefore, bias has a larger impact empirically as the horizon increases.

In order to relate to Equation 4.24, we can also examine MSE/T. For C-PCA the MSE/T is 11.2, 12.5, 16, and 29.4 at 1, 3, 12, and 48 months, respectively. For PCA it is 11.3, 12.7, 16.5, and 29.7. The pattern is clear in going from one to three to twelve months: MSE/T is getting progressively larger for PCA compared to C-PCA. Thus, it would seem that $\lambda_{PCA} > \lambda_{C-PCA}$, as expected from Equations 4.19 to 4.24. At the longest horizon of 48 months, we can infer (roughly) the lambdas. For C-PCA the lambda is roughly 1.9% and for PCA it is roughly 3.4%. The difference in lambdas between the two models is approximately the same for the other horizons as well, and the lambdas are what should be expected based on the bias for each model. We know that the larger lambda for PCA will make it perform worse as the horizon grows.

Note that this effect would be amplified further if the investor's loss function was asymmetrically shifted towards bias. Table 4.3 shows how a slightly higher weight on bias squared can make a large difference. This can be seen by comparing the magnitudes of the bias squared and variance terms (which together compose MSE). Bias squared is of an extremely small magnitude compared to variance, so placing a weight even slightly above an equal weight would dramatically change the resulting loss. The ratio of bias to RMSE can be examined to obtain the economic magnitude of how much more bias matters with a longer horizon. At one month, bias only makes up around 2.4% to 3.6% of RMSE depending on the model (and ignoring covariance terms in RMSE). However, this increases to 4.1%-5.8% at three months and 8%-10.9% at twelve months. At the longer-term horizon of 48 months, bias makes up as much as 18.5% of RMSE. It is expected that this would increase further as the horizon grows beyond four years (which we find in unreported results).

4.4.5 Rolling Window

The same out-of-sample analysis is performed, but with a rolling 60-month window rather than an expanding window with an anchored starting point. This is done both for robustness and because many researchers use a rolling 60-month window. Table 4.4 shows that while both PCA and C-PCA easily beat the traditional asset pricing models, C-PCA has the lowest MSE. However, it is only slightly lower and PCA has a slightly lower bias. In comparing Table 4.4 to Panel C (so that the sample period is the same), it can be seen that using a rolling window results in lower MSE for all models. However, for C-PCA the bias is now slightly higher. It appears that a smaller, rolling window allows the model to capture the local variation, resulting in less variance and lower MSE. However, if one would rather have a lower bias, then a larger, expanding window would be the proper choice. Also, the benefits of imposing the APT restrictions are greater when using a larger estimation window. In both Panels B and C of Table 4.2, C-PCA is able to drastically reduce bias (while having either a slightly higher or lower MSE, depending on the sample). However, with the rolling window the bias is not reduced (although MSE is still slightly lower). Therefore, if an investor/practitioner is interested in obtaining a lower bias (perhaps due to their asymmetric loss function or time horizon), then she should use an expanding window.

[Table 4.4 goes here]

4.4.6 Using in-sample betas

Another test is performed to see how well the factors of each model explain future returns. Here the portfolio weights are applied to the next month's realized portfolio returns (so that the factor realization is stored each month based on information up to the previous month). Where the methodology differs is that now instead of applying the previously estimated betas, a full-sample regression (similar to following the Black, Jensen, and Scholes, 1972 approach) is run of portfolio returns on factor returns.[42] The sample here is the previous out-of-sample estimation period (1977 – 2015 for the expanding window). While this is not a true out-of-sample test in that it cannot be applied in real-time by a potential investor, it still gives an idea of how well these portfolio weights (and consequently factors) explain future returns. Additionally, the betas are now allowed to be constructed optimally based on the realized returns.

Table 4.5 shows the regression results for the 1977 – 2015 sample, where the expanding window is used. The results with and without a constant (alpha) are shown for both PCA and C-PCA. Note that these are averages of the thirty time series regressions. Now that we allow each portfolio to optimize its betas based on the factor realizations, the C-PCA appears to better explain returns. The R-squared values are higher than those for PCA factors, and the MSE is lower. Further, when allowing for a non-zero alpha, the C-PCA approach produces an

[42] We also perform the traditional Fama and MacBeth (1973) test on all models, but the results did not vary much across models and are therefore omitted, but available upon request.

alpha that is 4 basis points lower per month. It can also be seen that only the first factor for each model is priced, although all four help explain covariance risk.

[Table 4.5 goes here]

The same analysis is repeated with the rolling window, where now the sample period is 1932 – 2015. Table 4.6 shows that while the MSE is slightly higher and the R-squared is slightly lower for C-PCA, it does produce an average alpha that is 10 basis points lower per month. Again, it appears that when imposing the APT restrictions, it is better to use a larger window. PCA is slightly better when implementing a rolling 60-month window, consistent with the findings in Table 4.4. The short estimation window may not allow for sufficient information in generating the weights, which should hold one-month-ahead out of sample if expected returns are efficiently estimated.

[Table 4.6 goes here]

4.5. Conclusion

In this paper we derive an approach in which the APT result of zero pricing error is imposed on principal components analysis (PCA). We also show that PCA can be represented as an approach that minimizes the sum of squared errors of a panel (or maximizing first pass time series fit). Thus, our approach imposes a perfect second pass fit (or up to a specified fraction of mispricing) while simultaneously maximizing first-pass fit. While PCA has the lowest possible mean squared error (MSE), we examine the breakdown of MSE into variance and bias (squared) terms. In sample, PCA will have the lowest MSE (and variance) of any model, while C-PCA (the model imposing APT restrictions) will have the lowest bias if mispricing is set to zero. We argue that for most applications of asset pricing models (cost of capital, forecasting, hedging, etc.) the practitioner may prefer lower bias rather than variance/MSE. Relatedly, any practitioner with a more long-term horizon would most likely prefer a model that is closer to the target on average (less bias), rather than being concerned with period-to-period fluctuations around the target (variance). We also show both analytically and empirically that a model with a lower bias will perform relatively better in terms of MSE as the investment horizon grows.

Due to the in-sample properties of the two models, an out-of-sample test is needed in order to properly compare the two. We implement a hedging approach based on portfolio realizations, where the weights and betas based on previous in-sample estimation are applied to the realized portfolio returns in the following month. We use the thirty Fama-French industry portfolios due to their weak factor structure in order to give a stronger test. With in-sample estimation starting (and anchored) in 1927 and out-of-sample estimation starting in 1977 and ending in 2015, we show that bias is reduced by almost one-third when imposing the APT restrictions while MSE is only slightly higher (not significantly different). Both in sample and out of sample, an investor or practitioner could adopt our approach and greatly reduce bias while only sacrificing a bit of MSE. We also show that allowing betas to be optimized in-sample (with factor weights applied out-of-sample) results in C-PCA better explaining returns in terms of both bias and MSE. For robustness purposes we also examine a rolling 60-month window. However, it appears that a larger sample is needed to properly estimate the factors. We also show that PCA and C-PCA perform significantly better than the traditional asset pricing models (CAPM, Fama-French three-factor model, and Carhart model).

It is argued that in order to properly implement the APT model, the model restrictions should be imposed a priori. Doing so results in lower bias out-of-sample, which would be desirable in certain applications. The methodology is derived herein, with possible extensions examined as well. This allows for future research to implement our approach for various applications. We implement the model to a portfolio hedged against systematic risk so that there is no need to estimate the risk premia, and show that bias is indeed reduced greatly. Future research in this area should also consider the breakdown of MSE and whether bias or variance is more important in the appropriate context.

References

Bai, J., and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica, 70*(1), 191-221.

Bartholdy, J., and Peare, P. (2005). Estimation of expected return: CAPM vs. Fama and French. *International Review of Financial Analysis, 14*(4), 407-427.

Basu, S., and Markov, S. (2004). Loss function assumptions in rational expectations tests on financial analysts' earnings forecasts. *Journal of Accounting and Economics, 38*, 171-203.

Best, M. J., and Grauer, R. R. (1991). On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results. *Rev. Financ. Stud., 4*(2), 315-342.

Bhatia, R. (2013). *Matrix analysis* (Vol. 169). Springer Science and Business Media.

Black, F., Jensen, M. C., and Scholes, M. S. (1972). The capital asset pricing model: Some empirical tests.

Bruner, R. F., Eades, K. M., Harris, R. S., and Higgins, R. C. (1998). Best practices in estimating the cost of capital: survey and synthesis. *Financial Practice and Education, 8*, 13-28.

Bussière, M., Hoerova, M., and Klaus, B. (2015). Commonality in hedge fund returns: Driving factors and implications. *Journal of Banking and Finance, 54*, 266-280.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance, 52*(1), 57-82.

Chen, N.-F., Roll, R., and Ross, S. A. (1986). Economic Forces and the Stock Market. *The Journal of Business, 59*(3), 383-403.

Chen, S.-J., Hsieh, C.-H., and Jordan, B. D. (1997). Real Estate and the Arbitrage Pricing Theory: Macrovariables vs. Derived Factors. *Real Estate Economics, 25*(3), 506-523.

Chou, P.-H., Ho, P.-H., and Ko, K.-C. (2012). Do industries matter in explaining stock returns and asset-pricing anomalies? *Journal of Banking and Finance, 36*(2), 355-370.

Clatworthy, M. A., Peel, D. A., and Pope, P. F. (2012). Are analysts' loss functions asymmetric? *Journal of Forecasting, 31*(8), 736-756.

Connor, G., and Korajczyk, R. A. (1993). A Test for the Number of Factors in an Approximate Factor Model. *The Journal of Finance, 48*(4), 1263-1291.

Cummins, J. D., and Phillips, R. D. (2005). Estimating the Cost of Equity Capital for Property-Liability Insurers. *Journal of Risk and Insurance, 72*(3), 441-478.

DeMiguel, V., Garlappi, L., and Uppal, R. (2007). Optimal Versus Naive Diversification: How Inefficient is the 1/ N Portfolio Strategy? *Rev. Financ. Stud., 22*(5), 1915-1953.

Diebold, F. X., and Mariano, R. S. (2002). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics, 20*(1), 134-144.

Easton, P. D. (2004). PE Ratios, PEG Ratios, and Estimating the Implied Expected Rate of Return on Equity Capital. *The Accounting Review, 79*(1), 73-95.

Easton, P. D., and Monahan, S. J. (2005). An Evaluation of Accounting-Based Measures of Expected Returns. *The Accounting Review, 80*(2), 501-538.

Elliott, G., Komunjer, I., and Timmermann, A. (2008). Biases in macroeconomic forecasts: irrationality or asymmetric loss? *Journal of the European Economic Association, 6*(1), 122-157.

Fama, E. F., and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, 33*(1), 3-56.

Fama, E. F., and French, K. R. (1996). Multifactor Explanations of Asset Pricing Anomalies. *The Journal of Finance, 51*(1), 55-84.

Fama, E. F., and French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics, 43*(2), 153-193.

Fama, E. F., and MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy, 81*(3), 607-636.

Fan, K. (1949). On a Theorem of Weyl Concerning Eigenvalues of Linear Transformations I. *Proceedings of the National Academy of Sciences, 35*(11), 652-655.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The Generalized Dynamic-Factor Model: Identification and Estimation. *Review of Economics and Statistics*, *82*(4), 540-554.

Fuller, R. J., and Kerr, H. S. (1981). Estimating the Divisional Cost of Capital: An Analysis of the Pure-Play Technique. *The Journal of Finance, 36*(5), 997-1009.

Giacomini, R., and White, H. (2006). Tests of Conditional Predictive Ability. *Econometrica, 74*(6), 1545-1578.

Goldenberg, D. H., and Robin, A. J. (1991). The Arbitrage Pricing Theory and Cost-Of-Capital Estimation: The Case of Electric Utilities. *Journal of Financial Research*, *14*(3), 181-196.

Goyal, A., Perignon, C., and Villa, C. (2008). How common are common return factors across the NYSE and Nasdaq?. *Journal of Financial Economics, 90*(3), 252-271.

Grinblatt, M., and Titman, S. (1987). The Relation Between Mean-Variance Efficiency and Arbitrage Pricing. *The Journal of Business, 60*(1), 97-112.

Harvey, C. R., Liu, Y., and Zhu, H. (2015). … and the Cross-Section of Expected Returns. *Rev. Financ. Stud., 29*(1), 5-68.

Herskovic, B., Kelly, B., Lustig, H., and Van Nieuwerburgh, S. (2016). The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics, 119*(2), 249-283.

Hou, K., van Dijk, M. A., and Zhang, Y. (2012). The implied cost of capital: A new approach. *Journal of Accounting and Economics*, *53*(3), 504-526.

Huberman, G., Kandel, S., and Stambaugh, R. F. (1987). Mimicking Portfolios and Exact Arbitrage Pricing. *The Journal of Finance, 42*(1), 1-9.

Jolliffe, I. (2002). *Principal component analysis*. John Wiley and Sons, Ltd.

Jong, J.-C., and Kotz, S. (1999). On a Relation between Principal Components and Regression Analysis. *The American Statistician, 53*(4), 349-351.

Kan, R., Robotti, C., and Shanken, J. (2013). Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology. *The Journal of Finance*, *68*(6), 2617-2649.

Kan, R., and Smith, D. R. (2008). The Distribution of the Sample Minimum-Variance Frontier. *Management Science, 54*(7), 1364-1380.

Kan, R., Wang, X., and Zhou, G. (2016) Optimal Portfolio Selection with and without Risk-Free Asset. *SSRN Working Paper*.

Kan, R., and Zhou, G. (2007). Optimal Portfolio Choice with Parameter Uncertainty. *Journal of Financial and Quantitative Analysis, 42*(3), 621-656.

Kavussanos, M., Marcoulis, S., and Arkoulis, A. (2002). Macroeconomic factors and international industry returns. *RAFE, 12*(12), 923-931.

Koedijk, K. G., and van Dijk, M. A. (2004). Global Risk Factors and the Cost of Capital. *Financial Analysts Journal, 60*(2), 32-38.

Kozak, S., Nagel, S., and Santosh, S. (2014). Interpreting factor models. Working Paper

Lee, C., Ng, D., and Swaminathan, B. (2009). Testing International Asset Pricing Models Using Implied Costs of Capital. *Journal of Financial and Quantitative Analysis, 44*(2), 307-335.

Lehmann, B. N., and Modest, D. M. (1988). The empirical foundations of the arbitrage pricing theory. *Journal of Financial Economics, 21*(2), 213-254.

Leitch, G., and Tanner, J. E. (1991). Economic forecast evaluation: profits versus the conventional error measures. *The American Economic Review*, 81(3), 580-590.

Lewellen, J., Nagel, S., and Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics, 96*(2), 175-194.

Li, R.-C. (2015). Rayleigh Quotient Based Optimization Methods for Eigenvalue Problems. *Matrix Functions and Matrix Equations, 19*, 76-108.

Li, Y., Ng, D. T., and Swaminathan, B. (2013). Predicting market returns using aggregate implied cost of capital. *Journal of Financial Economics, 110*(2), 419-436.

Litzenberger, R., Ramaswamy, K., and Sosin, H. (1980). On the CAPM Approach to the Estimation of A Public Utility's Cost of Equity Capital. *The Journal of Finance, 35*(2), 369-383.

Ludvigson, S. C., and Ng, S. (2009). Macro Factors in Bond Risk Premia. *Rev. Financ. Stud., 22*(12), 5027-5067.

MacKinlay, C.A., and Pástor, Ľ. (2000). Asset Pricing Models: Implications for Expected Returns and Portfolio Selection. *Rev. Financ. Stud., 13*(4), 883-916.

Mallows, C. L. (1973). Some Comments on C p. *Technometrics, 15*(4), 661-675.

Michaud, R. O. (1989). The Markowitz Optimization Enigma: Is 'Optimized' Optimal? *Financial Analysts Journal, 45*(1), 31-42.

Nardari, F., and Scruggs, J. T. (2007). Bayesian Analysis of Linear Factor Models with Latent Factors, Multivariate Stochastic Volatility, and APT Pricing Restrictions. *Journal of Financial and Quantitative Analysis, 42*(4), 857-891.

Patton, A. J., and Timmermann, A. (2007). Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics, 140*(2), 884-918.

Pratt, S. P., and Grabowski, R. J. (2008). *Cost of capital*. John Wiley and Sons.

Pukthuanthong, K., and Roll, R. (2014) A Protocol for Factor Identification. *SSRN Working Paper*.

Pástor, Ľ., Sinha, M., and Swaminathan, B. (2008). Estimating the Intertemporal Risk-Return Tradeoff Using the Implied Cost of Capital. *The Journal of Finance, 63*(6), 2859-2897.

Roll, R. (1977). A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. *Journal of Financial Economics, 4*(2), 129-176.

Roll, R., and Ross, S. A. (1980). An Empirical Investigation of the Arbitrage Pricing Theory. *The Journal of Finance, 35*(5), 1073-1103.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory, 13*(3), 341-360.

Simin, T. (2008). The Poor Predictive Performance of Asset Pricing Models. *Journal of Financial and Quantitative Analysis, 43*(2), 355-380.

Stock, J. H., and Watson, M. W. (2002). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association, 97*(460), 1167-1179.

Table 4.1: Bai and Ng (2002) results

Results are shown for the Bai and Ng (2002) criteria, using the 30 Fama-French industry portfolios with a sample of January 1927 - December 2015. The top panel shows the loss function results for each criterion and k factors from one to five. The bottom panel shows the optimum number of factors based on each criterion. The three different panel information criteria (IC) and panel C_p criteria (PC) are shown, and k* gives the optimal number of factors which gives the lowest loss function. See Bai and Ng (2002) for details.

k=	IC1	IC2	IC3	PC1	PC2	PC3
0	-5.2909	-5.2909	-5.2909	0.00504	0.00504	0.00504
1	-6.2904	-6.2894	-6.2926	0.00176	0.00176	0.00176
2	-6.3479	-6.346	-6.3524	0.00161	0.00161	0.00161
3	-6.3713	-6.3685	-6.378	0.00154	0.00155	0.00154
4	**-6.3733**	**-6.3695**	**-6.3823**	**0.00152**	**0.00152**	0.00151
5	-6.3664	-6.3616	-6.3775	0.00152	0.00153	**0.00151**

	k*
IC1	4
IC2	4
IC3	4
PC1	4
PC2	4
PC3	5

TABLE 4.2: In-Sample and Out-Of-Sample Performance

Panel A shows the full, in-sample results for the 30 Fama-French industry portfolios for the various models/approaches. The sample is January 1927 - December 2015. Panel B shows the one-month-ahead out-of-sample results with an anchored starting point and initial 50-year estimation window. The sample is January 1927 - December 2015, with out-of-sample hedging starting in January 1977 and ending in December 2015. Panel C shows the same OOS results, but with an initial 60-month window, with hedging starting in January 1932, ending again in December 2015. All panels show the mean squared error (MSE) and bias (absolute value of the average pricing error), averaged across all 30 portfolios. Bias is multiplied by 100 and reported in monthly percent form. MSE is multiplied by 10,000 to be consistent. The error term comes from directly applying weights and betas calculated from the particular approach to give a predicted return for each month (error is difference between realized return and predicted return). Note that MSE and Bias is not calculated via regression. The 2nd pass fit in Panel A is the adjusted R-squared value of a regression of mean returns regressed on 1st pass coefficients. The 2nd pass fit in Panels B and C is the adjusted R-squared value of a regression of mean returns regressed on mean predicted returns. Error ratio is the sum of squared errors (mean returns less mean predicted returns, squared) over the sum of squared mean returns, multiplied by 100. PCA is principal components analysis, C-PCA is the constrained PCA introduced in this paper, CAPM is capital asset pricing model, FF3 is the Fama-French three-factor model, and Carhart adds the momentum factor to the FF3 model. Results are shown for the case with zero mispricing (γ=0) and k=4 (four factors).

Panel A: In-Sample Results

	PCA	C-PCA	CAPM	FF3	Carhart
MSE	10.75	11.09	18.03	16.52	16.40
Bias	0.11	0	0.14	0.18	0.15
2nd pass fit	-0.78	0.41	-1.51	-0.78	-0.42

Panel B: Out-Of-Sample Results Starting in 1977

	PCA	C-PCA	CAPM	FF3	Carhart
MSE	11.35	11.45	18.30	17.72	17.72
Bias	0.15	0.11	0.19	0.21	0.19
2nd pass fit	-0.09	0.44	-0.90	-1.14	-0.77
error ratio	8.15	4.20	14.24	16.06	13.20

TABLE 4.2 continued

	Panel C: Out-Of-Sample Results Starting In 1932				
	PCA	C-PCA	CAPM	FF3	Carhart
MSE	11.29	11.23	17.71	16.76	16.85
Bias	0.12	0.08	0.15	0.17	0.14
2nd pass fit	-0.82	0.00	-2.07	-2.93	-1.66
error ratio	3.27	1.79	5.74	7.07	4.87

Table 4.3: Varying Time Horizon Out of Sample

This table shows the mean squared error (MSE), root mean squared error (RMSE), bias/absolute value of the average error, bias squared (bias sq), variance of the error (var), and standard deviation of the error (std dev) for the constrained principal components analysis (C-PCA) and principal components analysis (PCA) models with the 30 Fama/French industry portfolios as the test assets. Also shown is the proportion of bias to RMSE (multiplied by 100) in order to give an economic interpretation of how much bias contributes to the loss. The sample is January 1927 - December 2015 (out-of-sample estimation ending at December 2015). Bias, standard deviation, and RMSE are multiplied by 100 and reported in percent form. MSE, bias squared, and variance are multiplied by 10,000 to be consistent. Bias is reported as monthly returns. The error term comes from directly applying weights and betas calculated from the particular approach to give a predicted return for each month (error is difference between realized return and predicted return). Note that MSE and Bias is not calculated via regression. Each panel shows the results when varying the forecast/hedging horizon from 1 month to 3, 12, and 48 months. The in-sample estimation window starts at January 1927, and out-of-sample estimation starts at January 1932 giving an initial estimation window of 60 months, which is expanding. Monthly returns are used to find the loadings in all cases.

1-month returns

Model:	MSE	RMSE	Bias	Bias sq	Var	Std dev	Bias/RMSE
C-PCA	11.23	3.35	0.08	0.01	11.22	3.35	2.39
PCA	11.29	3.36	0.12	0.01	11.28	3.36	3.57

3-month returns

Model:	MSE	RMSE	Bias	Bias sq	Var	Std dev	Bias/RMSE
C-PCA	37.59	6.13	0.25	0.06	37.53	6.13	4.08
PCA	38.02	6.17	0.36	0.13	37.89	6.16	5.84

12-month returns

Model:	MSE	RMSE	Bias	Bias sq	Var	Std dev	Bias/RMSE
C-PCA	192.18	13.86	1.11	1.23	190.95	13.82	8.01
PCA	197.97	14.07	1.53	2.34	195.63	13.99	10.87

Table 4.3 continued

48-month returns

Model:	MSE	RMSE	Bias	Bias sq	Var	Std dev	Bias/RMSE
C-PCA	1413.42	37.60	5.20	27.04	1386.38	37.23	13.83
PCA	1424.19	37.74	6.97	48.58	1375.61	37.09	18.47

TABLE 4.4: Out-Of-Sample Performance with a Rolling Window

This table shows the one-month-ahead out-of-sample results when using a rolling 60-month window for the 30 Fama-French industry portfolios for the various models/approaches. The sample is January 1927 - December 2015, with hedging starting in January 1932 and ending in December 2015. The above results show the mean squared error multiplied by 10,000 (MSE) and bias (absolute value of the average pricing error) multiplied by 100, averaged across all 30 portfolios. The error term comes from directly applying weights and betas calculated from the particular approach's in-sample estimation to give a one-month-ahead predicted return for each month (error is difference between realized return and predicted return). Note that MSE and Bias is not calculated via regression. 2nd pass fit is the adjusted R-squared value of a regression of mean returns regressed on mean predicted returns. Error ratio is the sum of squared errors (mean returns less mean predicted returns, squared) to the sum of squared mean returns (reported as multiplied by 100). PCA is principal components analysis, C-PCA is the constrained PCA introduced in this paper, CAPM is capital asset pricing model, FF3 is the Fama-French 3-factor model, and Carhart adds the momentum factor to the FF3 model. Results are shown for the case with zero mispricing (γ=0) and k=4 (four factors).

	PCA	C-PCA	CAPM	FF3	Carhart
MSE	10.12	10.05	17.04	15.90	16.15
Bias	0.09	0.10	0.12	0.16	0.13
2nd pass fit	-0.01	-0.05	-0.89	-2.14	-1.30
error ratio	1.82	1.91	4.02	5.75	4.19

Table 4.5: Out-Of-Sample Factor Weights with In-Sample Betas

This table shows the results the regression results of returns on one-month-ahead factor returns. Reported are the mean for the 30 Fama-French industry portfolios. To form the one-month-ahead factor realizations, the factor weights are found in-sample and the one-month-ahead factor realization is stored based on those weights. The in-sample estimation period for the factor weights starts in January 1927, utilizing an expanding window with an anchored starting point. The regression sample is January 1977 to December 2015. F1 refers to the first factor, F2 the second, etc. Reported are coefficients and t-stats in parentheses. Also reported are the adjusted R-squared values and mean-squared error of the regression, multiplied by 10,000 (mean of the 30 portfolios). Results are shown for the case with zero mispricing (γ=0) and k=4 (four factors).

	PCA		C-PCA	
constant	0.0018 (0.23)		0.0014 (0.11)	
F1	0.1810 (27.27)	0.1812 (27.58)	0.1817 (24.09)	0.1818 (25.05)
F2	0.0092 (0.91)	0.0095 (0.92)	-0.0081 (0.32)	-0.0070 (0.37)
F3	0.0052 (0.13)	0.0055 (0.14)	-0.0085 (-1.11)	-0.0087 (-1.12)
F4	0.0099 (0.36)	0.0103 (0.38)	-0.0015 (-0.50)	-0.0009 (-0.44)
adj. R-squared	0.63	0.63	0.68	0.68
MSE	14.12	14.18	11.89	11.86

Table 4.6: Out-Of-Sample Factor Weights with In-Sample Betas (Rolling Window)

This table shows the results the regression results of returns on one-month-ahead factor returns. Reported are the mean for the 30 Fama-French industry portfolios. To form the one-month-ahead factor realizations, the factor weights are found in-sample and the one-month-ahead factor realization is stored based on those weights. The in-sample estimation period for the factor weights starts in January 1927, utilizing a rolling 60-month window. The regression sample is January 1932 to December 2015. F1 refers to the first factor, F2 the second, etc. Reported are coefficients and t-stats in parentheses. Also reported are the adjusted R-squared values and mean-squared error of the regression, multiplied by 10,000 (mean of the 30 portfolios). Results are shown for the case with zero mispricing (γ=0) and k=4 (four factors).

	PCA		C-PCA	
constant	0.0081 (4.01)		0.0071 (3.50)	
F1	-0.0536 (-8.03)	-0.0532 (-7.91)	0.0191 (1.81)	0.0227 (2.11)
F2	0.0109 (0.50)	0.0123 (0.56)	-0.0452 (-2.77)	-0.0422 (-2.55)
F3	-0.0286 (-0.86)	-0.0362 (-1.11)	-0.0699 (-2.54)	-0.0729 (-2.70)
F4	-0.0851 (-2.33)	-0.0923 (-2.53)	-0.0348 (-0.98)	-0.0376 (-1.06)
adj. R-squared	0.07	0.05	0.06	0.04
MSE	43.64	44.30	44.25	44.76

Appendix 4.A

```
%================================================================
MATLAB sample program designed to optimally choose k factors
(portfolios chosen from among n zero-investment assets) to
minimize the unexplained 'first-pass' errors subject to a
targeted fraction of 'second pass' pricing errors (astar). Output
provides Q as k vectors of portfolio weights normalized to square
to one. Further a and s provide the fractions of pricing errors
and unexplained variance, respectively, whereas a0 and s0 provide
the benchmark, unrestricted fractions of these.
===============================================================

clear all; clc;

%User Inputs k and astar:
k=2; %chosen number of factors; must be less than n
astar=0.00;  %chosen fraction of pricing errors (sum of squared
alphas relative to sum of squared mean returns).
%astar should logically be between 0 and 1, but also less than
a0, the implied a for lambda=0 for the optimization to make sense
(otherwise the constraint is not binding).

load FF25.txt;
FF=FF25(:,2:26);  %Fama-French excess returns for 25 size and
value sorted portfolios starting July 1963 ending June 2005
[t,n]=size(FF);

Sigma=cov(FF); %covariance matrix for the N asset returns
mu=mean(FF)'; %vector of mean returns over the T periods of the N
assets
lambda=0; %lambda is the lagrangian multiplier constraint for the
pricing errors constraint.  Initial value is zero.
eps=0.0000000001; %tolerance level for deviations from the target
fraction of squared pricing errors.
diff=-5; %the experimental change in the lagrangian multiplier
a=0.10; %initial level for the fraction of pricing errors
its=0; %iterations counter
CHECK=[]; %storing results for all iterations

%Loop to converge to true lambda (and Q) contingent on choice of
astar
while (a-astar)^2>eps && its<1000; %continue revisions until
pricing errors are within tolerance or the number of iterations
becomes too large
V=Sigma+lambda*mu*mu'; %V is the matrix for which eigenvectors
are found
[Q,Delta]=eigs(V,k); %'eigs' easier than 'eig' since it provides
Qk
abar=a; %abar accounts for the lagged iteration of a
a=(mu'*mu - (mu'*Q*Q'*mu))/(mu'*mu); %defines fraction of pricing
errors
```

```
s=1-trace(Q'*Sigma*Q)/trace(Sigma); %defines fraction of
unexplained variance
if (a-astar)*(a-abar)>0   %if a worsens reverse the change in
lambda
    diff=-diff/2; %reverse change in lambda but at half the pace
end
lambda=lambda+diff;  %update lambda in the direction that brings
pricing errors closer to target
its=its+1;
CHECK=[CHECK;[lambda a s]];
end

lambda=lambda-diff %corrected for the unnecessary change in
lambda in the final iteration.
%Note that lambda must be positive. If not, change the sign or
lower the value of the initial 'diff' choice.
lamcheck=trace((Delta-Q'*Sigma*Q)/((1-a)*mu'*mu))  %check that
lambda is indeed the equilibrium value

[Q0, Delta0]=eigs(Sigma,k); %find eigenvectors for unconstrained
case
a0=(mu'*mu - (mu'*Q0*Q0'*mu))/(mu'*mu); %provides fraction
pricing errors in unconstrained case.
s0=1-trace(Q0'*Sigma*Q0)/trace(Sigma);
%Check to make sure that a0>astar

Q,[a,a0;s, s0] %factor weights (Q) plus fractions of pricing
errors and unexplained variance in optimized case (column 1) and
benchmark case (column 2).
```

Appendix 4.B

Consider the following optimization problem:

$$(4.B1) \quad \underset{\mathbf{Q},\lambda}{Min} \quad \left[E(\mathbf{e}'\mathbf{e}) + \lambda(\boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha} - \boldsymbol{\gamma}\,\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})\right],$$

where $0 \le \gamma < 1$ represents the target level for the sum of squared alphas relative to the sum of squared mean returns. Using equations (4.1) and (4.2), the properties of traces, and taking expectations

$$(4.B2) \quad \underset{\mathbf{Q},\lambda}{Min} \quad Tr\left[(\mathbf{I}-\mathbf{BQ}')\,\Sigma\,(\mathbf{I}-\mathbf{QB}')\right]$$

$$+\lambda\left[\mu'(\mathbf{I}-\mathbf{BQ}')\,\Sigma^{-1}\,(\mathbf{I}-\mathbf{QB}')\mu - \gamma\,\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}\right]$$

The $\mathbf{B}$ are set exogenously equal to the standard OLS betas of equation (4.4):

$$(4.B3) \quad \mathbf{B} = \boldsymbol{\Sigma}\mathbf{Q}[\mathbf{Q}'\boldsymbol{\Sigma}\mathbf{Q}]^{-1},$$

Using the definition of $\mathbf{B}$ and the properties of traces in (4.B2), dropping constants and multiplying by -1 gives

$$(4.B4) \quad \underset{\mathbf{Q}}{Max} \quad Tr\left([\mathbf{Q}'\boldsymbol{\Sigma}^2\mathbf{Q}][\mathbf{Q}'\boldsymbol{\Sigma}\mathbf{Q}]^{-1} + \lambda\boldsymbol{\mu}\boldsymbol{\mu}'[\mathbf{Q}(\mathbf{Q}'\boldsymbol{\Sigma}\mathbf{Q})^{-1}\mathbf{Q}']\right),$$

and in addition λ is set to force $\boldsymbol{\alpha}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha} = \boldsymbol{\gamma}\,\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$. Using Trace properties:

$$(4.B5) \quad \underset{\mathbf{Q}}{Max} \quad Tr\left([\mathbf{Q}'(\boldsymbol{\Sigma}^2 + \lambda\boldsymbol{\mu}\boldsymbol{\mu}')\mathbf{Q}][\mathbf{Q}'\boldsymbol{\Sigma}\mathbf{Q}]^{-1}\right).$$

This is equivalent to

(4.B6) $$\underset{\mathbf{X}}{Max} \quad Tr\left([\mathbf{X}'(\mathbf{\Sigma}+\lambda\mathbf{mm}')\mathbf{X}]+\mathbf{\Delta}(\mathbf{I}-\mathbf{X}'\mathbf{X})\right),$$

where $\mathbf{X}=\mathbf{\Sigma}^{1/2}\mathbf{Q}$, $\mathbf{m}=\mathbf{\Sigma}^{-1/2}\boldsymbol{\mu}$ and $\mathbf{X}'\mathbf{X}=\mathbf{Q}'\mathbf{\Sigma}\mathbf{Q}=\mathbf{I}$ and $\mathbf{\Delta}$ is the multiplier enforcing this constraint. Then the first-order conditions are

(4.B7) $$(\mathbf{\Sigma}+\lambda\,\mathbf{mm}')\mathbf{X}=\mathbf{X}\,\mathbf{\Delta}.$$

(4.B8) $$\mathbf{V}(\lambda)\,\mathbf{X}=(\mathbf{\Sigma}+\lambda\,\mathbf{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{\Sigma}^{-\frac{1}{2}})\mathbf{X}=\mathbf{X}\,\mathbf{\Delta}$$

Pre-multiplying (4.B7) by $\mathbf{X}'$ gives

(4.B9) $$\mathbf{\Delta}=\mathbf{X}'\mathbf{\Sigma}\mathbf{X}+\lambda\,\mathbf{X}'\mathbf{mm}'\mathbf{X}$$

and subsequently pre- and post-multiplying both sides by $\mu_k=\boldsymbol{\mu}'\mathbf{Q}=\mathbf{m}'\mathbf{X}$ and its transpose, respectively, and using the normalization that $\mathbf{X}'\mathbf{X}=\mathbf{I}$ implies for the optimal choice of $\mathbf{X}$ that $\boldsymbol{\mu}_k'(\mathbf{\Delta}-\mathbf{X}'\mathbf{\Sigma}\mathbf{X})\boldsymbol{\mu}_k=\lambda(\mathbf{m}'\mathbf{XX}'\mathbf{m})^2$. To impose $\boldsymbol{\alpha}'\mathbf{\Sigma}^{-1}\boldsymbol{\alpha}=\boldsymbol{\gamma}\,\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}=\gamma\mathbf{m}'\mathbf{m}$ eliminate $\boldsymbol{\alpha}$ from equation (4.2). This produces $\mathbf{m}'\mathbf{XX}'\mathbf{m}=\boldsymbol{\mu}'\mathbf{QQ}'\boldsymbol{\mu}=(1-\gamma)\,\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}=(1-\gamma)\,\mathbf{m}'\mathbf{m}$. Then, $\mathbf{m}'\mathbf{XX}'\mathbf{m}=(1-\gamma)\,\mathbf{m}'\mathbf{m}$ and $\boldsymbol{\mu}_k'(\mathbf{\Delta}-\mathbf{X}'\mathbf{\Sigma}\mathbf{X})\boldsymbol{\mu}_k=\lambda(\mathbf{m}'\mathbf{XX}'\mathbf{m})^2$ jointly imply

(4.B10) $$\lambda=\boldsymbol{\mu}_k'(\mathbf{\Delta}-\mathbf{X}'\mathbf{\Sigma}\mathbf{X})\boldsymbol{\mu}_k\,/(1-\gamma)^2(\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu})^2.$$

Thus we use (4.B8) and (4.B10) together with $\mathbf{X}=\mathbf{\Sigma}^{1/2}\mathbf{Q}$, $\mathbf{X}'\mathbf{X}=\mathbf{Q}'\mathbf{\Sigma}\mathbf{Q}=\mathbf{I}$, and $\mu_k=\boldsymbol{\mu}'\mathbf{Q}$ to obtain the appropriate portfolios given OLS betas.

Chapter Five: Conclusion

This dissertation provides evidence that returns are predictable, enough so to justify acting on the forecasts (which is consistent with more recent literature but is a relatively new development). However, this is a relatively new development. The equity premium can be predicted so that market timing raises average returns. Individual portfolios can also be predicted out of sample with an appropriate asset pricing model. Information can be gleaned from these forecasts. For example, since size and book-to-market sorted portfolios can predict the equity premium (as shown in Chapter Three), size and value stocks may have higher returns because they reflect future investment opportunities. New asset pricing factors, and the approach for deriving them, are introduced in this dissertation as well. These new factors allow for better out-of-sample estimation of industry portfolio returns.

The first essay shows that investor sentiment can drive market mispricing, which is followed by market correction. This negative relationship with future excess market returns allows for investor sentiment to predict market returns. A new measure of investor sentiment is introduced, which is designed to remove economic fundamentals/rationality from an existing market-based measure of sentiment. The methodology removes a composite index of fundamentals to avoid over-fitting.[43] The results show that as more fundamentals are removed, the

[43] This is another over-arching theme of the dissertation: summarizing a large panel of variables/predictors avoids over-fitting and allows for better estimation. The other benefit is that the single index can parsimoniously summarize a large set of data.

negative relationship with future returns becomes stronger. This validates the noise trader hypothesis, while also being one of the few studies to show that sentiment can have a market-wide impact on returns. Aside from introducing a new way to measure sentiment, this essay also contributes to the field of study by showing that sentiment has significant forecast power. The forecasts could provide a potential investor with significantly positive abnormal returns, while also beating other benchmarks and competing models/sentiment measures.

The second essay's main contribution is to provide a different specific mechanism to explain the size and value premiums. The essay also shows that utilizing a Partial Least Squares approach applied to a large cross-section of portfolio returns can significantly predict the equity premium out of sample. A composite index is formed from portfolio returns (where portfolios are formed based on size and book-to-market sorts), which can be viewed as a separate risk factor/ICAPM state variable. Size and value portfolios can predict future excess market returns, so they reflect changes in future investment opportunities. Further, this new risk factor has potential applications which are explored. The factor portfolio itself could be invested in, which is found to actually have a higher Sharpe ratio than the market. Also, the factor can be used to hedge against future market risk, so a risk-hedging model is examined. Risk is reduced and Sharpe ratios are improved based on an optimization approach applied out of sample. Finally, the second essay offers a potential explanation for why an equal-weighted diversification strategy typically outperforms other more advanced approaches out of sample: the equal-weighted approach loads more heavily on size and value stocks, which are found to have a higher price of risk.

The third essay contributes to the field by providing a new approach to estimating asset pricing factors which perfectly explain mean returns in sample. The approach places a constraint on principal component analysis that forces zero alphas (mispricing). This essentially forces that the arbitrage pricing theory holds perfectly, and then maximizes first pass fit subject to that constraint. Thus, the approach is able to completely eliminate bias in-sample, while also maximizing the explained covariance among the test assets. It is shown that only a bit of mean squared error is sacrificed when imposing this restriction. The essay explores the potential benefits and applications of an unbiased model. An important contribution of the essay, though, is the out-of-sample estimation with these new factors. The factors (through their weights and the betas of individual portfolios) greatly reduce bias when used to hedge out of sample, compared to the traditional asset pricing models and the unconstrained principal component analysis. Further, the mean squared error is either the lowest of all models or only slightly higher than principal components analysis.

Overall, this research emphasizes the importance of out-of-sample estimation and return predictability. The traditional asset pricing models can explain cross-sectional or mean returns, but what matters in practice is having a good estimation of expected returns. This dissertation explores different approaches to moving out of sample. One way is to use the arbitrage pricing theory, forming factors that perfectly explain mean returns as in the model. This allows portfolio returns to be hedged out of sample with a reduced bias. The first two essays estimate the equity premium out of sample. The second essay shows that a composite index of portfolio returns can predict the equity premium, which allows for a bridge to an existing asset pricing model, the ICAPM. The first essay shows that behavioral finance has a place in predicting market returns. Specifically, investor sentiment has a market-wide impact that presents a predictable pattern of market returns.

This research should be of interest not only to the academic community but to the investor/practitioner community as well. In this light, potential applications are performed in each essay that could be implemented in real time. The results show that each essay's approach allows for beneficial out-of-sample estimation. In most cases the forecasts are both statistically and economically significant. While return predictability has started to gain acceptance in the academic community, market timing still has a negative connotation. This dissertation attempts to some arguments in support of market timing although it is never fully possible to distinguish unobserved risk premia and abnormal returns. Lastly, this dissertation attempts to bridge the gap from traditional asset pricing where cross-sectional or mean returns are explained to out-of-sample estimation and forecasting.